Vol. 60, No. 4 July, 1953

Psychological Review

CARROLL C. PRATT, Editor
PRINCETON UNIVERSITY
Lorraine Bouthilet, Managing Editor

CONTENTS

The Size-Distance Invariance Hypothesis.....F. P. KILPATRICK & W. H. ITTELSON 223

Formalization of Language Systems for Behavior Theory.....F. H. GEORGE 232

Concepts as Operators.....JOE ADAMS 241

Expectancies and Hullian Theory.....RICHARD A. BEHAN 252

Stimulation Satiation: An Explanation of Spontaneous Alternation and Related Phenomena.....MURRAY GLANZER 257

The Backward Curve: A Method for the Study of Learning.....KEITH J. HAYES 269

A Theory of Stimulus Variability in Learning.....W. K. ESTES & C. J. BURKE 276

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

The *Psychological Review* is devoted primarily to articles in the field of general and theoretical psychology. This area is obviously difficult to define and delimit, but in view of the large number of manuscripts sent to the editor on all kinds of topics an attempt has to be made to draw the line somewhere.

Ordinarily manuscripts that run to more than about 7500 words are not accepted. This policy is followed partly in an effort to reduce lag of publication and partly from the conviction that brevity which is not inconsistent with clarity is the best way to present an argument.

If an author is prepared to pay for the cost of publishing his article, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors.

Tables, footnotes, and references as well as text of manuscripts should be typed double-spaced throughout.

Editorial communications regarding manuscripts to be printed during 1953 should be sent to the present editor, Carroll C. Pratt, Princeton University. *All new manuscripts should henceforth be sent to the incoming editor,* Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan. The following persons have agreed to serve as editorial consultants: Solomon Asch, Robert Blake, Stuart W. Cook, Clyde Coombs, Leon Festinger, D. O. Hebb, Harry Helson, Carl I. Hovland, E. Lowell Kelly, David Krech, Robert W. Leeper, Robert B. MacLeod, David C. McClelland, G. A. Miller, Gardner Murphy, Oscar Oeser, Carroll C. Pratt, David Shakow, Eliot Stellar, S. S. Stevens, Eric Trist, Edward Walker, Robert White.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.
$6.50 volume $1.25 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1940

VOL. 60, NO. 4 JULY, 1953

THE PSYCHOLOGICAL REVIEW

THE SIZE–DISTANCE INVARIANCE HYPOTHESIS

F. P. KILPATRICK AND W. H. ITTELSON

Princeton University

Extremely common in the literature of visual perception are explanations or descriptions of the relationship between size, distance, and visual angle which resort to the use of mathematical expressions which are special cases of the proportions existing in similar triangles. Such applications of Euclidean plane geometry to psychological relationships, with their consequent, often uncritical, mixture of physical and psychological variables, offer attractively simple solutions. However, the value of these solutions rests at bottom on the degree to which they actually are consistent with the observed relationships. Consequently, a careful scrutiny in relation to experimental evidence of the hypothesis from which such mathematical expressions are derived would seem to be in order.

The basic expression follows directly from an elementary theorem of plane geometry. The proportions between corresponding parts of similar triangles shown in Fig. 1 can be expressed as

$$\frac{a_1}{b_1} = \frac{a_2}{b_2}.$$

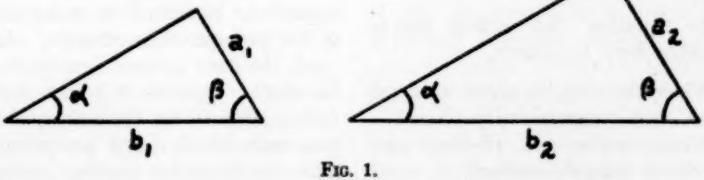

FIG. 1.

It follows directly from this that specifying a (Fig. 2) determines the relationship between a, b, and β. Specifically this relationship is given by

$$\tan \alpha = \frac{a \sin \beta}{b - a \cos \beta}. \qquad [1]$$

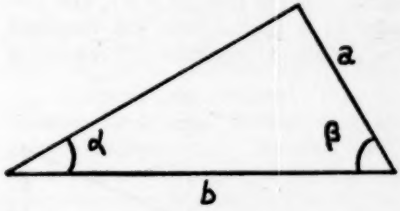

FIG. 2.

We can apply this simple geometry to physical objects in physical space, if we confine ourselves to sizes and distances which are neither too small nor too large.

Let S be the physical size of an object (Fig. 3) at a physical distance D from some reference point O, let a be the angle subtended by S at the point O, and let δ be the deviation of S from the vertical, or the *inclination* of S.

Equation [1] becomes

$$\tan \alpha = \frac{S \cos \delta}{D - S \sin \delta}. \qquad [2]$$

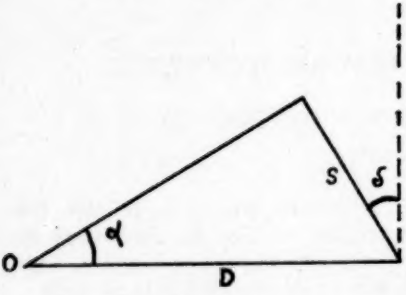

FIG. 3.

If we limit ourselves to angles sufficiently small so that $\tan \alpha = \alpha$, and consider as a special case the condition $\delta = O$ (Fig. 4), equation [2] reduces to

$$\alpha = \frac{S}{D}. \qquad [3]$$

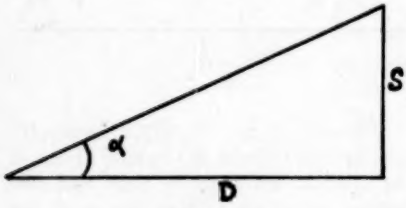

FIG. 4.

In this paper we shall consider the implications only of equation [3], bearing in mind at all times that it is only a special case of equation [2], which is in turn a special case of the three-dimensional relationship, including the most general case in which S is a curved surface.

From equation [3] certain simple conclusions may be drawn.

a. When the angle is given, size and distance vary proportionally.

b. When angle and distance are given, size is uniquely defined.

c. When angle and size are given, distance is uniquely defined.

d. When size is given, angle and distance vary inversely.

e. When distance is given, angle and size vary proportionally.

This simple geometry of the right triangle is of interest to the student of space perception because the apparent sizes and apparent distances of objects seem to conform to similar or identical rules. Indeed, in many cases equation [3] seems quite accurately to describe the relationship between apparent size, apparent distance, and visual angle. This fact has led some psychologists to apply this equation to the description of psychological events. In many instances, the resulting equations injudiciously mingle physical and psychological quantities. An example is

$$\text{visual angle} = \frac{\text{apparent size}}{\text{physical distance}}.$$

Another familiar example which leans more heavily on psychological quantities is

$$\text{retinal size} = K \frac{\text{apparent size}}{\text{apparent distance}}.$$

However stated, one assumption is central to all such applications of simple geometry to perceived space: *a retinal projection or visual angle of given size determines a unique ratio of apparent size to apparent distance.* This is the Size-Distance Invariance Hypothesis.

This hypothesis has been utilized by psychologists in two ways. First, as *explanation,* some form of invariance hypothesis is central to many theories of the perceptual constancies. As thus used, the most general statement of the invariance hypothesis holds that constancy depends on the existence of two processes which vary reciprocally so that the resultant product always re-

mains the same. This type of explanation as related to the size-distance problem was first proposed by Wheatstone; it is central in the Gestalt theory of constancy, implied in much of Boring's writing, and again crucial and explicit in Gibson's theory, to mention a few important examples (cf. the discussion in [11]). Also, it has found expression in recent issues of this journal in the writings of Schlosberg (16), who employed it to explain experiments in the visual perception of size and distance, including a number of the Ames demonstrations, and of Gelinsky (8), who used it as a basis for a derivation and quantitative formulation of "a general unifying law of visual space perception . . ." (p. 475).

In this paper we shall not be directly concerned with the use of the invariance hypothesis as an explanatory concept in perceptual theorizing, but rather with its second major use, that of *description*. We choose to concentrate on this latter aspect because it is more fundamental; the adequacy of a formulation as explanation depends directly on its adequacy as description. Is the invariance hypothesis adequate to describe the observed relationships between apparent size, apparent distance, and visual angle? The existing evidence will be examined and a new experiment reported in an effort to answer this question.

That the size-distance invariance hypothesis does describe this relationship in many instances is well attested. No attempt will be made here systematically to survey the literature on this subject but rather simply to indicate the general nature of the supporting evidence. Three cases may profitably be distinguished.

a. Visual angle constant. The invariance hypothesis demands that apparent size and apparent distance vary proportionally, and indeed this is the most commonly reported result. Duke-Elder has summarized current evidence of this case by noting that "If we estimate the distance to be much greater than it is, we consider that if the object at this distance is so big, it must be very big indeed. Conversely, if an object seems nearer than it is, it seems to be smaller than it actually is" (6, p. 1071). This case is also the one with which Emmert's law is concerned, whose usual interpretation is in support of the invariance hypothesis.

b. Apparent distance constant. The invariance hypothesis demands that apparent size vary directly with visual angle, and again this is the most commonly reported result. This case is, in fact, closely related to the finding of most size-constancy work, that the proper estimation of size is dependent upon the proper estimation of distance.

c. Apparent size constant. The invariance hypothesis demands that apparent distance vary inversely with visual angle. This general finding has been repeatedly verified in a variety of cases.

It is clear, then, that considerable experimental evidence can be mustered in support of the invariance hypothesis, and indeed most writers on the problem of the relation between apparent size and apparent distance would seem to be in general agreement with this hypothesis. The prevailing view can be summarized in Sanford's words written half a century ago, ". . . size and distance are mutually determining. If the apparent distance is constant, the apparent size of the objects changes directly with the size of the retinal image; while if the apparent size is constant, the apparent distance changes inversely with the image. These are facts of very common observation" (15, p. 198).

Possibly the case could be allowed to rest there, were it not for the fact that there is also considerable experimental

evidence in support of the proposition that no aspect of perception can be a simple and invariant function of a few variables. This proposition is derived from the more general theory that the perceptual process is one which always and necessarily includes all aspects of the functioning of the organism (1, 4, 5, 11, 12, 13). To find that the size-distance relationship is an invariant one as so commonly reported would constitute an important exception to that theory. Consequently, it would appear to be of considerable systematic importance to note that the findings with respect to the invariance hypothesis are not entirely one-sided. Although the contradictory evidence is scattered and possibly not conclusive, it stems from a variety of experimental approaches.

a. Varying accommodation while looking at a fixed object. The relation between accommodation and apparent size under these conditions is well established. Anyone can verify for himself the truth of Hering's report that "when changing from a monocular fixation of a distant object to viewing one close by, the distant object appears to diminish in size. Conversely the size of the nearer object increases when accommodating for the farther" (10, p. 172). The relation of this effect to apparent distance is by no means so clearly established. This problem has never been systematically investigated, but it has been mentioned by many writers, who, although they are not in complete agreement, would seem to support von Kries in stating:

> When we gaze at any object with one eye screened, it is easy to notice that its apparent size varies with the state of accommodation. Every exertion of accommodation is accompanied by an apparent reduction in size, and every relaxation by an apparent magnification. The simplest explanation of this well-known and easily observed phenomenon would be to suppose that as the result of accommodation for near vision the object appears to be at the distance that ordinarily corresponds to this accommodation, that is, too near; and hence, as long as it subtended the same visual angle, its absolute size would appear to be less than it was. The trouble about this explanation is that, as Donders long ago rightly pointed out as something remarkable, this is not what actually happens. The object does appear to become smaller when accommodation is exerted, but it by no means appears to come nearer at the same time; on the contrary, it appears to recede. Thus, with a constant visual angle, we see here apparent size and apparent distance vary in the opposite sense (17, p. 389).

This is clearly and explicitly in contradiction to the invariance hypothesis.

It should, however, be pointed out that this evidence, as well as that in the following section, is complicated by the possibility that the physical size of the retinal projection varies with changes of accommodation. (See, for example, discussions in Neumueller [14] and Duke-Elder [6]).

b. Accommodative micropsia due to partial paralysis of accommodation. Partial paralysis of accommodation, most commonly by means of atropine, is accompanied by a reduction in apparent size, or micropsia. While this fact is well established, the relation to apparent distance again is not completely clear. Duke-Elder implicitly attributes this effect to the invariance hypothesis when he writes that ". . . in partial paralysis of accommodation . . . we volitionally expend a greater effort to see an object distinctly, . . . and we thus imagine the object to be quite near; but since the retinal image is of the same size, we think the object must be smaller than it is" (6, p. 1072). Hering, however, reported findings of Aubert which specifically refute this explanation. "Aubert after instillation

of 1/500 grain atropine, saw letters with the atropinized eye at a distance of 20 feet, apparently half the size as seen with the other. . . . In spite of this, strange to say, the apparently smaller objects did not seem nearer but farther than in the sound eye" (**10**, p. 172). This report is clearly in contradiction to the invariance hypothesis.

c. Visual tau effect. "If three discrete points of light are successively presented, and the distance between the points of light is equal, but the time interval between the first and second points of light is greater (lesser) than the time interval between the second and third, then the perceived space between the first and second will be greater (lesser) than the perceived space between the second and third points of light" (**7**, p. 483). Contrary to what one might expect in terms of the invariance hypothesis, there is no evidence that this well-attested phenomenon is accompanied by any apparent distance change. Of course, it might be argued that the perceived space between the successively presented lights is not "size" in the usual sense, as the lights which set the "space limits" do not stimulate the retina simultaneously. Nevertheless, pending further experimentation, the visual tau effect might reasonably be considered an exception to the invariance hypothesis.

d. Subjective factors. Some recent evidence seems to indicate that the apparent size of an object may be influenced by subjective factors such as the importance of the object to the observer (cf. discussions in [2, 3]). While such evidence is as yet inconclusive, if it should be substantiated, it would constitute an important exception to the invariance hypothesis.

e. Statistical considerations. Much if not all of the evidence which tends to support the invariance hypothesis is statistical in nature. That is, it states that the invariance hypothesis describes the mean of a large number of observations in a large number of cases. But as usually expressed, the invariance hypothesis is not a statistical statement; it is an equation which, within the limits of accuracy of measurement, must apply to the individual case. Therefore, even if the invariance hypothesis always adequately described the mean performance (which, as we have seen, it probably does not), it would still have to be modified drastically before it would be acceptable as a description of psychological events.

It is clear, then, that there exists a body of evidence which, although it approaches the problem obliquely, is yet relevant and contradictory to the invariance hypothesis. The evidence at the very least suggests that this hypothesis may not be adequate to describe the relationship between apparent size, apparent distance, and visual angle in all cases. There is need, nevertheless, for an experiment specifically designed to test the applicability and generality of the invariance hypothesis. In the experiment reported below visual angle was kept constant, apparent distance was influenced by over-lay indications, and apparent size by the size of the background. Reports were obtained of changes in apparent size and apparent distance and were analyzed as supporting or contradicting the invariance hypothesis.

DESCRIPTION OF THE EXPERIMENT[1]

The apparatus employed has been described and illustrated by the authors in a previous publication (**13**). Briefly, it consisted of a trapezoidal shape cut out of ⅛-in. aluminum, 39 in. long, 47¼ in. and 25 in. high at the long and the short ends, respectively.

[1] The authors are indebted to Charles T. Albert for collecting the data in this experiment.

Three rows of five openings somewhat like the panes in a window were cut out of the trapezoidal shape, and on it were painted shadows like those commonly seen on an ordinary window.

This trapezoid was suspended from the ceiling by two wires so that the sides were vertical and the center of the frame 46 in. above the floor. A black thread passed through the middle opening of the shorter side and was attached to four pulleys so that the thread was horizontal to the floor, 48 in. above the floor, and at right angles to the line of sight. A motor attached to one of the pulleys drove this thread at the rate of about half an inch per second. This apparatus was viewed from a distance of 10 ft. utilizing a headrest which permitted monocular observation only, and which was adjustable so that eye level could be maintained at 46 in. above the floor.

Black drapes were hung behind the trapezoid and between the trapezoid and the observer. In this latter drape, 4 ft. from the viewing point, a rectangular opening was cut through which the trapezoid could be observed. Behind this drape two 15-w. lights in photographic reflectors were mounted so that they gave balanced illumination to both sides of the trapezoid. With the exception of these two lights the room was completely dark.

Two objects were provided which could be attached to the black thread and thereby moved through the trapezoid. These objects were an ordinary playing card (three of diamonds, 2¼ in. × 3½ in.) and a piece of cotton approximately 2 in. in diameter.

Observers were twenty-four undergraduates at Princeton University, none of whom had any previous knowledge of the nature or purpose of this experiment. All of them had, or were corrected to, 20/20 visual acuity.

The trapezoid described above appears to *O* to be a rectangle with the actually far edge of the trapezoid appearing as the near edge of the rectangle and vice versa. Under these conditions, objects carried through the trapezoid in a straight path by the moving thread appear typically to move through an "S"-shaped path in the horizontal plane (see **13**). The problem, as related to the study of the invariance hypothesis, is to determine whether this change in apparent distance as the object moves is accompanied by the corresponding change in apparent size which would be predicted from the invariance hypothesis. In order to secure this information, each *O* was presented with three different situations.

1. *Trapezoid alone.* The *O* viewed the trapezoid and was asked to describe its orientation in space and its approximate dimensions.

2. *Trapezoid with moving objects.* Each of the two objects described above was viewed moving through the trapezoid from right to left and also from left to right. The order of presentation was randomized. The *O* viewed each condition twice and was asked to describe any changes in apparent size or distance of the moving object.

3. *Trapezoid with stationary objects.* Two playing cards were suspended from the stationary wire with their inner edges lined up from the viewing point with the outside edges of the trapezoid.

FINDINGS

1. *Trapezoid alone.* All *O*s described the trapezoid as a rectangle tipped sharply sideways with its left edge near and its right edge far. Estimates of its dimensions ranged between 1½ ft. × 2 ft. and 5 ft. × 3 ft., with 18 out of the 24 *S*s estimating the horizontal dimension to be at least 3 ft.

2. *Trapezoid with moving objects.* In 75 of the 96 trials, definite movement in depth of the test object was

reported, 3 ft. or more in 39 of the trials and less than 2 ft. in only 12 of the trials. Perceived movement in depth was about equally common for the two test objects; in the case of the playing card, such movement was seen in 36 out of 48 trials, while with the cotton it was seen in 39 out of 48 trials. No significant effect of the right-left or left-right orders appeared for either test object, although depth movement was reported a little more often for both objects in the right-left trials. These depth changes, when seen, were quite consistent in length throughout the trials of an individual, and, in general, corresponded to *O*'s earlier estimate of the horizontal dimension of the frame. There can be little doubt, then, that most *O*s in the majority of their trials did see the test object change distance, and that these changes were of sufficient magnitude relative to the viewing distance of 10 ft. that a corresponding change in size should have been perceived in all such cases if the invariance hypothesis were to be substantiated.

a. Playing card. In only 14 of the 37 trials where either size or distance change was seen could the *O*s' reports be interpreted to be in line with the invariance hypothesis. These 14 reports were that when the card approached, it became smaller, and when it moved away, it became larger. In the remaining 23 trials, three quite different size-distance perceptions were reported, all of them distinguished by their lack of accord with the invariance hypothesis. The most common of these deviant observations, reported in 15 trials, was that a marked change in distance was seen without any apparent alteration in size. Another report, occurring in 7 instances, was exactly the reverse of the invariance-hypothesis formulation; when the card approached it became larger, when it receded it became smaller. In the remaining instance no change in distance was reported but the *O* said he saw the card change size.

TABLE 1

REPORTED SIZE-DISTANCE PERCEPTIONS OF THE MOVING OBJECTS

(24 observers, 4 trials each)

Response Category	Number of Trials		
	Card	Cotton	Total
No change in size or distance	11	9	20
Changed in accord with invariance hypothesis	14	15	29
Changed, but *not* in accord with invariance hypotheses*	23	24	47
Total trials	48	48	96

* This category includes three types of size-distance relationships: (*a*) change of size and not distance, (*b*) change of distance and not size, and (*c*) decrease of size with increase in distance or vice versa.

b. Cotton. The results with the moving piece of cotton were substantially the same as those reported for the playing card. The numerical distribution of the findings is reported in the second column of Table 1.

3. *Trapezoid with stationary objects.* Here, also, with the visual angles subtended by the two playing cards the

TABLE 2

REPORTED SIZE-DISTANCE PERCEPTIONS IN THE STATIC SITUATION, RIGHT-HAND CARD STATED IN RELATION TO LEFT

(24 observers, 1 trial each)

Response Category	Number of Trials
Same size and distance	5
In accord with invariance hypothesis:	
Larger and farther	10
Not in accord with invariance hypothesis:	
Larger and nearer	5
Larger and same distance	2
Vacillating between "larger and farther" and "larger and same distance"	2
Total trials	24

same, there is no invariant relationship between reported apparent size and distance. The modal category of response (10 *Os*) is in line with the invariance hypothesis, but almost an equal number of *Os* (9 cases) reported perceptions which could not be so interpreted. The results of this part of the experiment are summarized in Table 2.

DISCUSSION

The size-distance invariance hypothesis states that for a given visual angle there is a unique and constant ratio of apparent size to apparent distance. The results reported above, together with the earlier evidence, clearly indicate that this hypothesis is not adequate to describe this relationship in all cases. The invariance hypothesis therefore loses its status as a powerful explanatory concept and becomes rather a description of results obtained under conditions which have yet completely to be specified.

This conclusion certainly should not be surprising. The invariance hypothesis is commonly expressed in the form

$$\text{visual angle} = \frac{\text{apparent size}}{\text{apparent distance}} \quad [4]$$

or some equation reducible to this. Such equations represent an attempt to take a relation derived under the axioms of Euclidean plane geometry and to transfer it bodily to the description of psychological events. There is no a priori reason why this should be possible.

When applied to physical objects in physical space, and within certain limiting conditions, the equation

$$\alpha = \frac{\text{physical size}}{\text{physical distance}} \quad [5]$$

always holds. This equation is useful to the psychologist because it defines the physical limits of the experience of the organism. Whenever the retina has been stimulated along a length describable by the visual angle α, this has always been produced by physical objects at physical distances uniquely specified by the above equation (or more generally by equation [2]). There have never been any exceptions to this rule; there never will be. When equation [4] adequately describes the reported perceptions, this simply means that the perceptions most nearly approximate the physical situations which have been related to all the size-distance experience of the organism, and which can be summarized by equation [5].

When behavior of this type is encountered, it can be labeled *invariance behavior*. The conditions under which invariance behavior can be observed have yet to be determined, but should provide a more satisfactory base line for evaluating experimental results than the usual criterion of constancy. It is, so to speak, what one would expect the organism to do if there were no reason for doing otherwise.

REFERENCES

1. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
2. BLAKE, R. R., & RAMSEY, G. V. (Eds.) *Perception, an approach to personality*. New York: Ronald, 1951.
3. BRUNER, J. S., & KRECH, D. (Eds.). *Perception and personality*. Durham: Duke Univer. Press, 1949.
4. CANTRIL, H. *The "why" of man's experience*. New York: Macmillan, 1950.
5. CANTRIL, H., AMES, A. JR., HASTORF, A. H., & ITTELSON, W. H. Psychology and scientific research. *Science*, 1949, **110**, 461–464, 491–497, 517–522.
6. DUKE-ELDER, W. S. *Textbook of ophthalmology*. St. Louis: Mosby, 1936.
7. GELDREICH, E. W. A lecture-room demonstration of the visual tau effect. *Amer. J. Psychol.*, 1934, **46**, 483–485.
8. GELINSKY, A. A. Perceived size and distance in visual space. *Psychol. Rev.*, 1951, **58**, 460–482.

9. HELSON, H., & KING, S. M. The tau effect: an example of psychological relativity. *J. exp. Psychol.*, 1931, **14**, 202–217.
10. HERING, E. *Spatial sense and movements of the eye.* (Trans. by C. A. Radde.) Baltimore: American Academy of Optometry, 1942.
11. ITTELSON, W. H. The constancies in perceptual theory. *Psychol. Rev.*, 1951, **58**, 285–294.
12. KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952.
13. KILPATRICK, F. P., & ITTELSON, W. H. Three demonstrations involving the visual perception of movement. *J. exp. Psychol.*, 1951, **42**, 394–402.
14. NEUMUELLER, J. The correction lens. *Amer. J. Optom.*, 1948, **25**, 247–261.
15. SANFORD, E. E. *A course in experimental psychology.* Boston: Heath, 1901.
16. SCHLOSBERG, H. A note on depth perception, size constancy, and related topics. *Psychol. Rev.*, 1950, **57**, 314–317.
17. VON KRIES, J. Notes. In H. von Helmholtz, *Physiological optics.* (Trans. by J. P. C. Southall.) Optical Society of America, 1925.

[MS. received April 10, 1952]

Psychological Review
Vol. 60, No. 4, 1953

FORMALIZATION OF LANGUAGE SYSTEMS FOR BEHAVIOR THEORY

F. H. GEORGE

University of Bristol

This paper is intended to further the analysis that has already been partially developed (6). It is now hoped to suggest steps of a more positive nature designed to develop adequate language systems for the discussion of behavior theories. It seems to the writer that two of the most pressing needs in modern behavior theory are: (*a*) The development of a more precise language for the discussion of the theory, since much of existing behavior theory is vitiated by an apparent inability to formulate sensible questions. The questions asked prior to experiment, for example, or those which form the basis of theory construction, are, all too often, either meaningless, circular, or insufficiently precise, and this is partly, at least, a function of the language used. (*b*) The development of systematic theories of behavior in molecular terms, as well as in molar terms, appears necessary to remove the vagueness that surrounds molar constructs. This vagueness can be greatly reduced by breaking the constructs down to a molecular level. This paper is wholly concerned with finding a solution to (*a*).

The Development of Formalization

"Formalization brings out clearly to what a large extent problems in science are not empirical questions but questions of linguistic convention. This also helps to eliminate useless controversy." This quotation from a memorandum by Woodger[1] forms the basis for this discussion. Formalization is, broadly, the process of reducing our linguistic systems to more precise logical forms; axioms, definitions, and so on. In this paper it is intended to indicate the desirable development which formalization of behavior theory should take. This approach, it is hoped, will especially be brought to the notice of those narrow experimentalists who repeatedly call for experiment, and decry theory, and continue to stumble through the maze of science erroneously believing themselves to be dealing wholly with facts, and never with linguistics.

It is apparent that there are many symbolic logics of various forms, some of which have already been used in the psychological field, so far with somewhat limited success. Such systems may be premature in a subject which is not clearly developed as a science. Their compact shorthand, although in some form certainly desirable as an ultimate aim, needs to be translated back into longhand for many purposes. Thus our aim is a rigorization of language, or the setting up of a compromise between rigid postulational systems of a narrow kind, such as the propositional calculus, and the vague, inadequate language of ordinary discussion. The principal trouble in existing psychological languages is either (*a*) they sprawl, introduce enormous numbers of new terms, and are vague and lend themselves to meaningless formulations, or (*b*) they are impressively precise and impose a rigidity on our knowledge of behavior that leads to an arbitrariness which allows of no real prediction. More precisely, (*a*) they allow no prediction which is precise, or (*b*) allow precise prediction which is almost wholly false to facts.

[1] Unpublished manuscript.

Ideally it seems that we should have a hierarchy of languages which moves from the very general to the extremely particular and allows us to suit the precision of our language to the precision desired in the discussion. We talk vaguely in ordinary conversation, but become more precise as the dimensions of our topic become more precise, and ultimately we are forced back, as is necessary, to definitions and even to the assumptions underlying our theoretical system. It is certainly true that this last and highest approximation, to which we refer in cases of obscurity, could well be in mathematical terms, or some such rigorous form. This paper, then, will attempt to suggest the way we may fill the gap between our ordinary language and our logical calculi.

In an earlier paper (6) the vagueness accruing at certain points even in rigorous theory, i.e., in the use of logical constructs, undefined terms, philosophical directives, many-meaninged terms, etc. was noticed. This should be borne in mind throughout this paper and will be referred to from time to time.

THEORY CONSTRUCTION

In science, the search for adequate methods of theory construction continues, and psychological theory is especially interested in deriving some relatively precise methods. In this, and subsequent paragraphs, it is intended to outline some of the problems and one way in which they might be solved.

In the first instance we may choose to build our models in at least one of two ways. We start from our experience in either case, but we may find it convenient to assume the existence of a real external world behind experience, which needs to be observed and mapped on various levels of language. This is what Pragmatic Realism aims to do (5, 15). We may, on the other hand, start from experience, without making any assumptions about a real external world; thus our theoretical model comes to consist of observational propositions, and our vague and difficult formulations will be defined operationally. This last method is generally referred to as some brand of Phenomenalism, Positivism, or Operationism (1, 2, 3).

I would submit that there is no obvious reason for not accepting a mixed Operational-Realist system wherein we may regard the Realistic model as being generally adequate, and the use of an operational definition as a form of limiting test where disputes, difficulties, etc. occur—generally at the borders of our knowledge at any given time. The ability to use such operational definitions is a guide to the degree of validity of the theory. The question of setting up criteria for deciding what constitutes a valid operation is a matter for convention, as is the problem of "observation-in-principle." To recapitulate, we are asserting that concepts, or formulations, need not, in general, be defined in a strictly operational form, but that we may conveniently use a realistic model subject to an operational test, wherever possible, and then define our terms in a manner we shall call a contextual definition.

There are now further problems in theory construction to be settled.

"INTENSIONAL" AND "EXTENSIONAL" MODELS

The question now to be considered is the actual method of construction that is best suited to behavior theory. Let us first consider the classical postulational method which has proved to be so successful in mathematics and mathematical science. The theories of Groups, Matrices, Euclidean Geometry, and Boolean Algebra are well-known examples of postulational systems. (It

will be noted that these markedly successful cases are essentially analytic, or conceptual systems.) In psychology Hull has, of course, used a kind of postulational method in the statement of his theory of behavior (9).

The postulational technique itself is not the question which I primarily wish to discuss, as I intend to assume that in some form this is essential. The question is rather: Is the intensionally *defined* system adequate? To put it in other words, we may start from a set of undefined terms and from their derived postulates, and then introduce the various conceptual terms such as learning, perception, motivation, habit, habituation, adaptation, reinforcement, and so on. These terms are not necessarily all to be included in the theory, but some such terms, many-meaninged in character, must certainly arise. The many-meaninged, or multiordinal (having a different meaning on each level of language) types of term must, and will, arise in scientific theory; and they present the central problem. Now again to put the question: are our definitions to be intensional?—i.e., is a class definition adequate? May we, for example, talk of the class of reinforcers, as does Meehl (14)? The answer appears to be that the notion of class is often objectionable and the source of both vagueness and crude approximation. The alternative is an extensional definition, which involves the enumeration of properties, and no question of class definition, although the word "properties" also has its drawbacks, as has been pointed out by Woodger (16).

Let us now compare what I have called intensional and extensional definitions. If class definition is accepted, then we set up a class of some kind, subject to some class criteria; we categorize in fact, and in doing this, differences are overlooked at the expense of similarities. The method is compact and powerful. Consider, for example, the enormous power inherent in the "Boolean Algebra of Classes." The power of the method depends on our ability to categorize without too much distortion. Can this always be done? The answer must be: not always in subjects served by ordinary language without vagueness and arbitrariness. On the other hand, extensional definitions do not allow conciseness, convenient generalization, or a calculus; they do, however, gain in accuracy of statement. The trouble essentially surrounds the question of classification. As soon as we classify, we overlook differences.

The problem may be solved by a compromise. It is literally true that behavior theory needs the benefits of both systems. Thus extensional definitions may be thought of as limiting cases of intensional definitions, when the number of classes tends to infinity. The form of definition to be employed must therefore allow a degree of approximation, suiting the dimensions of the subject discussed. We shall propose, then, that our definitions should start from a glossary definition—a short statement of the common properties involved in the various uses of the term. Then on each occasion that the term is used, it will be contextually defined. The contextual definition may be explicit, and certainly it should be capable of explicit definition. If we now place this common glossary definition, and its set of contextual definitions, in a model of an infinity of levels of language, it will be seen that we have made explicit a simple, and powerful, device for building language systems. The glossary definition, at any level, may be taken as a class definition, if, for example, for quantification purposes, it was necessary to have a precise, if arbitrary, statement. This may be profitable on some dimensions

at some times. At the same time there will exist a general extensional background of levels of propositions with any degree of accuracy of definition possible. A simple example of this usage must suffice. At the level of molar behavior theory a central term is "learning." A considerable amount of discussion takes place without an agreed criterion for learning. The result is confusion, in many cases, not only because two people use different criteria, but because they both, themselves, use the term in various ways. In order to discuss "learning," a start could be made with a glossary definition such as the one from the monograph of Hilgard and Marquis (8), and each time the term was used it would be in this general sense, unless some explicit contextual definition was appended.

Science has, of course, progressed by bringing more and more under less and less, in the sense that certain postulates can be used to subserve more and more diverse and complex propositions. It is this generalization, simplification, and unification that makes science so powerful. However, the price paid by this advance can be too great if it leads to arbitrariness, and this appears to be the present position in behavior theory. Certainly it appears as yet to be impossible to encompass the great variety of behavior in a rigid postulational system. This, at any rate, is an assumption we shall make. Thus explicitly an attempt must be made to build an extensional system which might, as it were, be fixed and made exact, at any level, at any time. In practice this will only be useful if we can suit the fixing to the appropriate dimensions.

Now our criteria for theory construction may be summarized. We must state our undefined, or ostensively defined terms, and in giving our ostensive discussion we must, as far as possible, explicitly state our philosophical directives (i.e., those personal prejudices, etc. that we possess). Next, we must build a set of postulates which will contain words such as "learning," "perception," and so on. Avoidance of these particular terms only throws the burden on some other many-meaninged terms. What has been said does not mean that we are rejecting the rigid postulational methods as used by Hull, and his associates (9, 10), or that we are returning to straightforward general discursive methods such as have been generally used in the work of Hebb, Thorndike, Tolman, etc. The general discursive method is too vague and puts us in a position which leads to discussion about terms, and the linguistic conventions, which are not clear-cut, and this state of affairs seems quite inadequate. The alternative, however, is not at this stage to go to the strictest type of Hull system. This—and this is the principal point made in this article—leads to an arbitrariness which is too narrow for the present development of behavior theory.

The steps that could be taken in learning theory would be to reduce the number of undefined concepts by defining in terms of a small nucleus of undefined terms (necessary in any verbal system), and by setting up a glossary definition for vague, previously undefined concepts, many-meaninged terms, etc. where the "glossary definition" should refer to the common properties of all usages, and then a particular contextual definition given to each usage. If a term is used in, say, two ways that have no common property, then a new term should be introduced. If this pattern is followed explicitly, rather than implicitly, much clarity would result.

It is necessary to bear in mind the distinction made between propositions,

which are part of our strict behavior model, and propositions of the metalanguage, which refer to the propositions of the model. This distinction is implicit in our linguistic theory which is assumed to be on any of a possible infinity of levels. This sort of distinction (13) is necessary to cater for antimonies, and is analagous to the Russell "Theory of Types." With respect to the terminology which we may employ, it is, I believe, true that the Carnap axiom (4) holds and that each subject will develop its own terminology, and will gradually discard the terms that perform no useful function in the subject.

Thus it is proposed that the construction of behavior theory involves first the rigorization of our language systems. We then proceed extensionally, by careful contextual definition of terms, to theorems, which are derived from postulates which refer both to our explicitly stated undefined terms and our glossary definitions. It is worth mentioning that some such language system has recently been developed by Woodger (16) and previously a somewhat similar system was developed by Lesniewski. Woodger aims to avoid the enumeration of properties and the difficulties of class, and he has built an extensional type of calculus which allows discussion to take place and allows degree of accuracy to be involved by devices known as "time-slicing." This is similar to the methods of "indexing" suggested by Korzybski (13), and involves, at the crudest level, the distinction between Walter Scott 1843 and Walter Scott 1844. This can clearly allow of degrees of accuracy, analagous to the infinitesimal calculus, as 1843 can be subdivided into March 1843, March 3rd 1843, and so on indefinitely as we take smaller and smaller "time-slices." This is precisely the sort of graded accuracy on various levels of discourse, or levels of abstraction, that is needed in behavior theory, although we have used the phrase "enumeration of properties," to avoid the notion of definite class. It may now be seen that the precision of glossary and contextual definitions is essentially related to dimensions, i.e., what is adequate as a context definition on one level, may only serve as a glossary definition on another level.

Applications in Behavior Theory

Of all branches of behavior theory, theories of learning and perception are the most philosophically and semantically sophisticated. The fields of abnormal behavior, Personality, and so on, are in a much more primitive state of development and are far from the Carnapian state of throwing out the functionally useless terms. Thus the sort of linguistic systems suggested apply even more obviously to the latter fields, but I believe they are also enormously important to the former. The whole subject of redefinition of psychological terms could well fill many volumes.

The first question is how far may terminology be stretched. The best example extant is perhaps the terminology of the conditional reflex. Recently the Pavlovian terminology of the conditional stimulus, conditional response, unconditional stimulus, unconditional response, generalization, etc. has been extended from the original "meaning" (8, 12) to include more complicated behavior activities. This involves the extension of definitions of some of the terms, and has the advantage of bringing more and more forms of behavior under the same postulational system. However, this is only worth while if the amount of stretching necessary in the definitions is not too considerable; otherwise, it would seem more useful to introduce a new term, and utilize it un-

til such time as the integration may be effected without undue stretch. The actual test must be that of overlap referred to previously. In any event here we may clearly see the use of the contextual definition. If we are discussing generalization, it is important, granted that it has to some degree been stretched, to say precisely, in each instance, what we mean by generalization, and if necessary proceed back to the observations to be carried out to elicit the phenomenon we are naming.

The broader problem of whether conditioning terminology should be used, or not, in the wider context of Instrumental conditioning, in a laudable attempt to bring more of behavior under one concise terminology, can only be decided by attempting to verify the theorems that emanate from the assumptions, i.e., is it possible to carry out the plan without serious distortions and omissions?

Gibson (7) has carried out something of the desired type of analysis for the term "set" with the idea of finding some common meaning for the various usages. He found none which might serve for a glossary definition, but instead, unearthed ambiguities and contradictions. It is precisely this messy uncontrolled growth of terminology that ultimately leads to a lack of clarity and vagueness in the subject that makes prediction impossible. It is clearly essential that further surveys of the kind made by Gibson should be carried out in order to clear up the confusions of the past, and to introduce greater rigor in our formulations in the future.

It is worth noting, parenthetically, that there is another reason why vagueness enters the behavior theories so far formulated: the neglect of the genuine comparative method in psychology. Terms such as "set," "reinforcement," etc. are being stretched too far if made to refer to *all* organisms, without reference to contextual differences.

In Hull's system we find, in spite of the relatively precise, perhaps overprecise, nature of its formulations, vagueness occurring over such words (undefined terms of the system) as "response." The truth of many of the assertions of Hull's system would appear to depend on the precise interpretation given to this term in various contexts.

"Learning" itself is notoriously a term signifying a process which is vague. Note the sort of distinction made by Humphrey (11). He regards "learning" as a modification of behavior "towards a biologically useful end," but not modification of just as precise a nature towards an end that is not useful, e.g., a seagull learning to follow ships to get food, as opposed to a seagull learning always to fly directly away from ships and starving. This is surely an odd insistence and implies a fundamental confusion, expressible in the statement that learning is not so-and-so, but such-and-such (11, p. 105), which fails to lay emphasis on the essential point that we may define the word learning in any manner we please as long as we have common agreement about that usage. It is unnecessary to add that many different usages of "learning" do in fact exist, although their differences are not always, by any means, made explicit.

The last example we can briefly remark on is the distinction made by some psychologists between insight and trial-and-error types of learning. If learning is not equatable with performance, then it is not immediately obvious that sudden changes of performance are sufficient to build a terminological distinction, and in any event, such distinction needs to be formulated with precision to be useful to science, whereas in the past it appears merely to

refer to different philosophical directives. If this is a genuine nonverbal distinction, then it needs to be formulated in terms that allow some clear operational distinction. This appears typical of the lack of rigor in the existing state of much psychological theory.

MOLAR AND MOLECULAR

A word on the interpretation given to the words "molar" and "molecular" seems apposite as these words are indeed many meaninged.

two points on a continuum, at least in one sense. This distinction is important because propositions which are vague, nontestable, etc. on molar levels may be testable on molecular, and vice versa. This even applies to different molar and different molecular levels. The problem can be simply illustrated diagrammatically. Figure 1 is a representation of the essence of behavior where *S*s stand for stimuli, *R*s for responses and *O* for the organism—to be regarded essentially, of course, as in

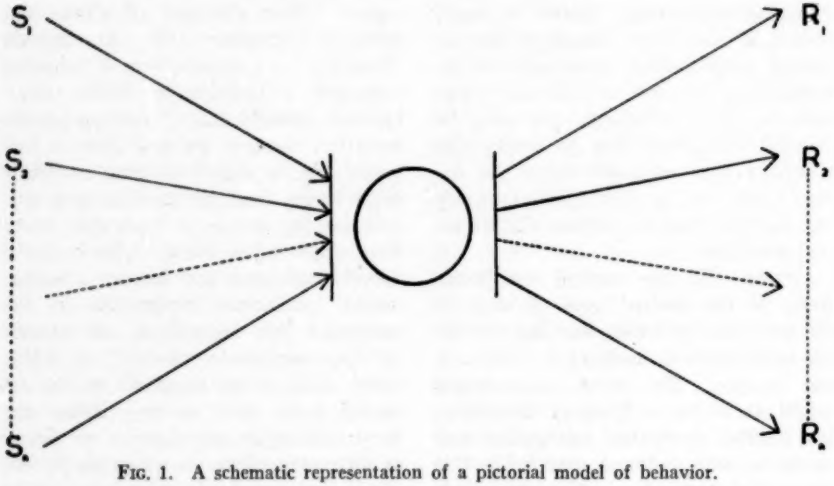

FIG. 1. A schematic representation of a pictorial model of behavior.

Granted that we see all of science on a simple plan of assumptions, consequences of assumptions, tests of consequences, etc.—going on indefinitely in a spiral, where we might equally well have written postulates, theorems, experiments to test theorems, postulates, etc. Granted also that we accept the need for precise propositions on some sort of a hierarchical model as has been suggested, with the added flexibility allowed by glossary and contextual definitions, then we should also expect to find that the molar-molecular apparent dichotomy really represents

an environment and in a space-time manifold. The molar theorist now acts as shown in Fig. 2. He attempts to describe behavior in terms of varied and complex stimulus and response conditions without carrying out tests, observations, etc. on the internal state of the organism. Thus the vagaries that attend making inferences about learning from performance are accounted for by logical constructs (these are represented by the "tags" in Fig. 2, strung together and involving any number). A consideration of latent learning, delayed reward, etc. from

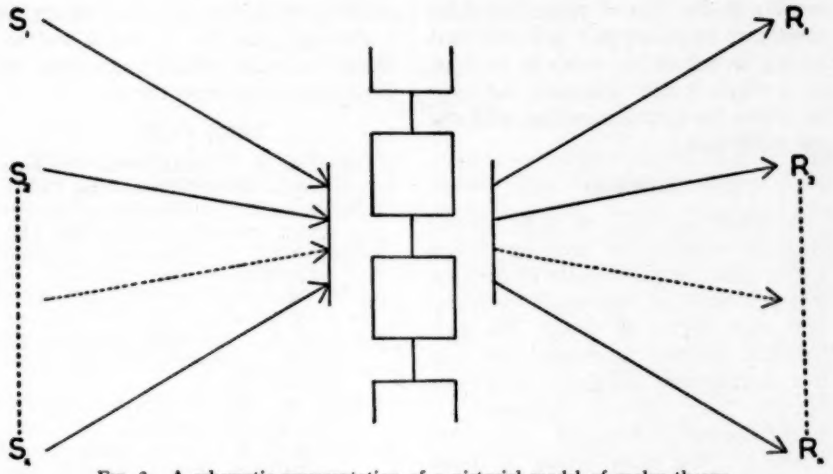

FIG. 2. A schematic representation of a pictorial model of molar theory.

molar theory itself is sufficient to show internal variation cannot always be neglected, even if this was not obvious from our knowledge of physiology, endocrinology, etc. Figure 3 illustrates the molecular viewpoint in essence. Here the molar constructs are broken down into molecular constructs where ⊞ represents diagrammatically the subdivisions of □. The great advance is simply that ⊞ can be regarded as part of □ and part of a molar theory and testable as such, and at the same time is testable on the neurological, endocrinological level. The double purpose is served of making explicit our molar constructs and opening up a further testing ground for their validity.

This whole matter is complex but the term molecular should be regarded

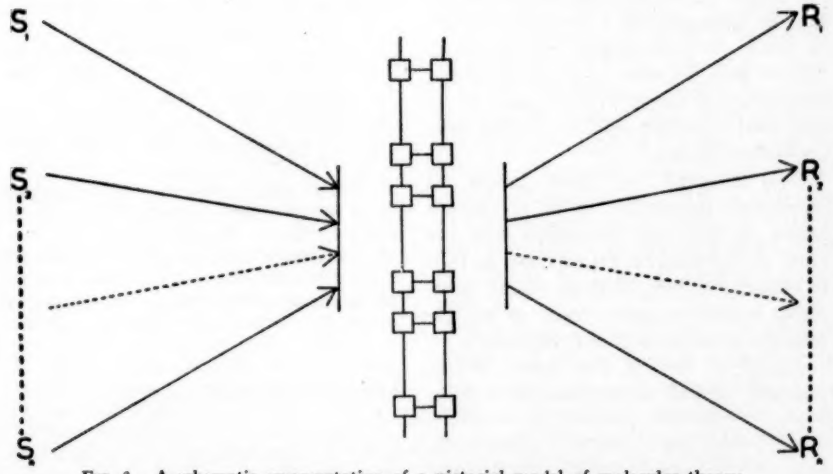

FIG. 3. A schematic representation of a pictorial model of molecular theory.

roughly in the light of tying our molar constructs to the organic data and thus hoping to reduce the work to be done by a single logical construct and thus to reduce the surplus meaning, and also the vagueness.

SUMMARY

It seems necessary to build precise language systems for behavior theory which allow for the breadth of the subject and which at the same time maintain some degree of rigor. The gaps between common languages and narrow postulational techniques need to be filled. The method suggested is straightforward, although sometimes exacting. We must accept a broad postulational form and be prepared to state our undefined terms, postulates, logical constructs, etc. and then allow for a nucleus of undefined terms that are ostensively defined and for "philosophical directives" which should be stated as fully as possible. Then we are left with theorems involving many-meaninged, multiordinal terms and complex undefined concepts, which are best treated by a glossary definition, supplemented by a contextual definition.

Further it will be recognized that we are, in language, on a possible infinity of levels of abstraction and we must differentiate between propositions about psychological theory (this paper is on this level) and propositions within psychological theory.

The attempt has been made to enumerate properties, or in Woodger's terms, to deal with the members of the class as individuals, rather than to talk of classes, as talk of class—more especially definite classes—leads to an unwelcome arbitrariness and vagueness.

Lastly, a few of the more obvious possible verbal confusions have been pointed out in psychological theory as a starting point for a verbal and semantic overhaul which seems vital to the future of behavior theory.

REFERENCES

1. AYER, A. J. *Language, truth and logic.* London: Oxford Univer. Press, 1936.
2. BRIDGMAN, P. W. *The logic of modern physics.* New York: Macmillan, 1927.
3. BRIDGMAN, P. W. The nature of some of our physical concepts. I, II, III. *Brit. J. Phil. Sci.*, 1951, **4**, 257–272; **5**, 25–44; **6**, 142–160.
4. CARNAP, R. Empiricism, semantics and ontology. *Revue internationale de Philosophie*, 1950, **11**.
5. FEIGL, H. Existential hypotheses. *Phil. Sci.*, 1950, **17**, 35–62.
6. GEORGE, F. H. Logical constructs and psychological theory. *Psychol. Rev.*, 1953, **60**, 1–6.
7. GIBSON, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.
8. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
9. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
10. HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
11. HUMPHREY, G. *The nature of learning.* London: Kegan Paul, 1933.
12. KONORSKI, J. *Conditioned reflexes and neuron organisation.* Cambridge: Cambridge Univer. Press, 1948.
13. KORZYBSKI, A. *Science and sanity.* (2nd Ed.) Lancaster, Pa.: Science Press, 1941.
14. MEEHL, P. E. On the circularity of the law of effect. *Psychol. Bull.*, 1950, **47**, 52–75.
15. SYMPOSIUM ON EXISTENTIAL HYPOTHESES. *Phil. Sci.*, 1950, **17**, 164–195.
16. WOODGER, J. H. Science without properties. *Brit. J. Phil. Sci.*, 1951, **2**, 193–216.

[MS. received April 3, 1952]

Psychological Review
Vol. 60, No. 4, 1953

CONCEPTS AS OPERATORS [1]

JOE ADAMS

Bryn Mawr College

It is well recognized that ordinarily, if not always, any object that is perceived is perceived either as a certain *kind* of object or as a *familiar* object or both. Almost every one agrees that most examples of this characteristic of perception imply some kind of trace or other prior existing process or organization which has been built up with past experience. Terms which have been used by psychologists to refer to these prior existing somethings are "concepts," "stereotypes," "schemata," "traces," "trace systems," "frames of reference," "enduring cognitive organizations," "cognitive maps," "belief value matrices," "conceptions," "categories," "meanings," etc.[2] A review of the literature employing these terms reveals two interesting points. First, there is not nearly as much overlap as one might expect. Psychologists using the term "concepts" have been interested almost exclusively in concept formation or development. Those using the terms "traces" or "trace systems" (14, 15, 16) have been concerned with the process of communication of the present stimulus with the trace or trace system (the Höffding function) or the communication of one trace with another. Psychologists using the terms "stereotypes," "schemata," or "frames of reference" (1, 2, 26, 27) have dealt with the functional aspects of these constructs in the perceptions, memories, thoughts, and behaviors of the individual. Possibly Krech and Crutchfield (19) with their discussion of cognitive organizations have gone furthest in the integration of these various aspects, but the research reviewed by Vinacke (33) and Leeper (21) illustrates how little integration of this kind has been achieved among psychologists in general.

A second point revealed by a review of the literature is the paucity of adequate terms and methods for the identification, description, and measurement of these constructs. Certainly concepts, etc. must have many characteristics and must reveal their existence in many different ways, but with a relatively few exceptions, psychologists write as though one NUZ is about like any other NUZ, one stereotype about like any other, and as far as investigations of what are called "concepts" go, methods are confined pretty much to sorting or labeling objects, and to formulating criteria, although in the case of "stereotypes" and "schemata" methods involving selective perception and distortions in memory have been employed.

The lack of adequate descriptive terms and methods may contribute to the striking discrepancies between the exciting hypotheses asserted by those who write "in the large" about cognitive processes or linguistic behaviors

[1] This paper, which is intended as a foundation for a forthcoming series of experimental papers, was written during the spring semester, 1952, while the author was a member of the Language and Symbolism project at the University of Michigan, sponsored by the Rockefeller Foundation. It is in part an outgrowth of study during a Social Science Research Council Fellowship, 1948–49, and in part an extension and revision of the author's unpublished Ph.D. dissertation, Princeton University, 1948. Grateful acknowledgment is made to Dr. Heinz Werner and Dr. Silvan Tomkins, who encouraged and advised the author during his SSRC Fellowship.

[2] Some psychologists prefer to deal with this phenomenon in a very different way and do not use these terms or any close equivalents.

and the actual results obtained by experimental research. Many psychologists have tended to ignore the descriptive and functional problems in favor of the genetic problem, i.e., here as in some other areas of psychology, psychologists have been much more interested in giving theoretical explanations of how something has been learned or derived from needs than in describing that something itself and demonstrating its ways of functioning. The history of psychology should give us a clue here. The study of perception made great strides forward when Katz and the gestalt psychologists emphasized the necessity of a description of perceptions rather than an "explaining away" of perception in terms of association or an artificial analysis of perception into hypothetical components, as the structuralists had done. Similarly, the study of behavior has been greatly aided by careful attention to an adequate description (usually in terms of accomplishment, e.g., pressing a bar, reaching a goal box) instead of an analysis into hypothetical components (e.g., reflexes in the Watsonian sense). Likewise, understanding of language is beginning to profit greatly by a careful description of utterances that actually occur (see, for example, the stimulating book by Fries [8]). It may well be that the psychology of cognitive processes will not advance much further until adequate descriptive methods are available.

The purpose of this paper is to propose a number of distinctions which can fruitfully be made in the description of concepts and which enable the use of a common terminology in the discussion not only of concept formation but also of the functioning of concepts in personality. These distinctions are also intended to imply a close relation between traditional psychological problems and more or less psychological hypotheses which have arisen in related fields, especially philosophy, sociology (the sociology of knowledge), and anthropology and linguistics (especially in the writings of Benjamin Whorf [37, 38]). Some indication will also be given of the procedures which are used in the actual application of the descriptive terms. For lack of space the full theoretical and applied value of these distinctions cannot be argued, but a few illustrations are given of how the terms apply to some current research. The terminology used in the paper is, it seems to the author, compatible with the usage of "concept" by psychologists in the fields of concept formation and development. On the other hand, it has been necessary to introduce a number of distinctions which either have not been made explicitly before by psychologists or, if made, have not been labeled. Several of these distinctions are very similar to distinctions made in logic and semantics,[3] and terms used by writers in these fields, especially Carnap (3), will be borrowed. The letter "c" will be prefixed to borrowed terms, to indicate that the usage, though similar, nevertheless has important differences.[4]

Suppose an individual perceives an object X as a certain kind of object, say A. Then we shall say that this individual has a concept C, and we can symbolize the perceptual event as $C(X) = A$. We shall call C an "operator" at the time it is being used to structure perception.

We shall also use the term "operator" for a concept that is being used in

[3] At least to the author these distinctions are similar. I do not agree with those who maintain that discussions of meaning in philosophy and in psychology have, or should have, nothing in common.

[4] The differences between the present usage and Carnap's usage arise from the fact that the terms in the present paper take into account the individual.

an analogous way in cognitive functions other than perception, in the strict sense of the term, such as what is sometimes called "social perception" (e.g., perceiving that Mr. F. likes Miss G., perceiving that Mrs. H. is insincere, etc.), and also in functions like remembering, thinking, imagining, conceiving. In the following discussion, the words "perception" or "perceiving" will be used for brevity, although the assertions are intended to cover other cognitive functions as well. The discussion assumes, therefore, as a working assumption, that the *same internal organizations that operate in one cognitive function operate in other functions.* This assumption is not unverifiable; it is already supported to some extent by available evidence, but if further research should demonstrate that it does not hold, there will be very little value in the notion of concept.

It should be emphasized that "concept" in the present discussion is not being used as a pure intervening variable (23). It is assumed that there are some kinds of internal organizations or processes which are responsible for categorical aspects of cognitive functions, and that these internal organizations have *properties which we must infer from observation.* Part of what follows will seem curiously empty or tautological if this is not kept in mind.

To refer again to our paradigm $C(X) = A$, the significance of the fact that A is not X, that in perceiving the world we use only a small number of the operators that might be used, and that different people, or even the same person at different times, may use different operators, has been discussed by so many psychologists, philosophers, sociologists, anthropologists, and linguists that no attempt will be made to discuss its significance here.[5] It may be pointed out, however, that the question of whether an individual *has* a concept (e.g., whether he can, under suitable instructions, make certain sortings or classifications and/or give a satisfactory definition) is probably not as important as the question of *under what conditions the concept will operate.*

A considerable amount of psychological research can be grouped under the following two problems:

1. What is the functional significance of individual differences in operators under certain standard conditions? Much of the research in projective techniques deals with this problem.

2. What are the factors (other than personality) making for the operation of one concept rather than another in the same individual? Under this heading would come much of the research on the effect of need, set, structural factors, and practice on perception. For example, Duncker's experiments on "functional fixedness of solution objects" (4) can be thought of as dealing with this problem, as he demonstrated that under certain conditions once an object has been operated on by a certain concept, it is more difficult for a different concept to operate on that particular object. Under other conditions, however, the operation of one concept greatly facilitates the operation of certain other concepts on the same object, as for example in racial or ethnic prejudice, where the perception of an individual as a member of a certain group reduces the number or strength of cues or information necessary to trigger the operation of concepts associated with that group. Prob-

[5] Among those who have remarked about the significance of this or related phenomena are Helmholtz, Wundt, James, Titchener, Piaget, Bartlett, Schachtel, Koffka, MacLeod, Sherif and Cantril, Allport and Postman, Krech and Crutchfield, Plato, Aristotle, Aquinas, Locke, Berkeley, Cassirer, De Laguna, Whitehead, Sapir, Whorf, Lady Welby, Lippmann, and Korzybski.

ably almost any operator increases the probability of operation of some concepts and decreases the probability of operation of others.

In our paradigm, $C(X) = A$, we call A a *category* of the concept C, and X an *argument*. We say then that a *concept operates upon an argument to yield a category*.[6] The concept C may have categories in addition to A. By the statement "B and A are both categories of C" we mean that whatever concept operates on an object X to yield A is the same concept as that operating on some object Y to yield B. There are methods for determining whether A and B are categories of the same C, but it will be easier to describe these once we have introduced a few additional terms.

We distinguish between the *category* A and the *verbal label* which the individual uses for the category or for a member of the category. We call such a label an "expressor" of the category A. A category may have more than one expressor. The expressors of a concept are the expressors of its categories.

A complete description of a concept includes two aspects, *content* and *formal properties*. We shall discuss each of these aspects in detail.

The most obvious way of describing concepts is in terms of the assignment of objects to the concept's various categories. In concept-formation experiments this is the most commonly employed test for determining whether a given concept has been formed; implicitly, therefore, this is the way the concept is described. To explicate this notion precisely we shall borrow some terms from logic and semantics. If $C(X) = A$, then we call X a "*c-denotatum*" of A (cf. Morris [24], Russell, J. S. Mill, etc.) and also of any of A's expressors. It is convenient to speak of the *c-extension with respect to a particular set of objects S*. If an individual selects from a set S all those which he is willing to classify as A's, then the latter subset is the c-extension of A with respect to S. The c-extension of a *concept*, on the other hand, is the *set of c-extensions of the categories of the operator;* it is what the mathematician calls a *partition* of a set of objects.

It should be noted that we have not said that any object that *has* certain properties is necessarily a c-denotatum of a certain expressor; it is at this point that we diverge from semantics into pragmatics. An object S is, in our usage, a c-denotatum of an expressor if and only if the individual whose concept we are studying applies that expressor, whether X actually has certain properties or not.

A description of a concept purely in terms of its extensions with respect to certain sets of objects leaves something out, namely what, in a certain sense, the individual "means" by his expressors.[7] We shall call the properties which the individual perceives[8] (or believes or conceives) the c-denotata as having the "c-intension" of the category (and its expressor). In some cases we can distinguish between the *defining c-intension* and the *incidental c-intension*, i.e., between those properties which *define* the category and those which tend to be *attributed* (perhaps with varying strengths of hypotheses)

[6] The present author likes to call these categories "*values*," as this usage is in accord with the logical analogue (an *operator* operates upon an *argument* to yield a *value*) and is more appropriate than "category" when the notion of a *nominal* operator, as discussed in this paper, is extended to that of an *ordinal, interval,* or *ratio* operator.

[7] The inadequacy of a description in terms of extensions is especially obvious in those cases in which the extension is null (e.g., "unicorn") or has only one member (e.g., "myself").

[8] The word "perceives" is used in the broadest sense, to include effects of cues of which the individual may not be conscious.

to any c-denotatum but are not defining properties.[9] For example, part of the defining c-intension of the expressor "Negro" by a prejudiced individual would probably be the property of having dark skin and part of the incidental c-intension might be the property of being innately stupid. *In most cases, however, a distinction between defining and incidental c-intensions will be very difficult, if not impossible, to make.* Even in scientific research these two parts of a c-intension are not always distinguished. If one reviews the history of the duplicity theory of vision, for example, one looks in vain for careful definitions of "rods" and "cones"; as a consequence of this ambiguity, many of the statements made in the development of the theory (including the general statement of the theory itself) are confusing mixtures of definitions, tautologies, and empirical assertions, and the extent to which these statements are verified by certain observations is left indeterminate. It is well recognized that in some cases in which this distinction *can* be made we categorize an object which clearly satisfies a *defining* c-intension, then assume without conscious inference that the *incidental* c-intension is also satisfied. Work on prejudice and stereotypes has emphasized this mechanism, discussed by psychologists under the term "assimilation" (Allport and Postman [1]), "frame of reference" (Sherif and Cantril [27]), or "incorporation into organized cognitive structures" (Krech and Crutchfield [19]).

We have defined "c-intension of a *category.*" The defining and incidental c-intensions of a *concept* are the sets of defining and incidental c-intensions of its categories.

Notions like that of c-intension are strongly rejected by many of those interested in problems of language and meaning. The argument runs thus: There are utterances (verbal behaviors), other behaviors, and objects; there may even be images, feelings, and other implicit responses, but look wherever you will, there are certainly no intensions of any kind, and to drag in some notion of properties being attributed to objects is only to introduce confusion, pseudo-problems, and a lot of fruitless verbiage. This argument does not convince the present author; to him the elimination of notions like c-intensions, though perhaps theoretically possible, greatly handicaps any attempt to deal with meaning and leads to rather cumbersome and peculiar circumlocutions.[10]

[9] The term "defining c-intension of the expressor," as here defined, has a meaning somewhat similar to that of Frege's *"Sinn"* (5), Carnap's "intension" (3), Morris's "significatum" (24), Lewis's "signification" (22), J. S. Mill's "connotation," etc., except that these terms are usually used by the authors with reference to languages or semantical systems instead of with reference to one individual's usage, or else they are used in a more restricted sense technically (e.g., cf. Morris). Carnap's "intension" includes the defining c-intension plus all its logical implications, which, in the language of the present paper, may or may not be part of the incidental c-intension. For example, the intension, in Carnap's sense, of the expressor "positive integer" would include the property of having a unique factorization into prime factors, as it has been proved that all positive integers do have unique factorizations, whereas this property would not for any individual's expressor be part of the defining c-intension, and would be a part of the incidental c-intension if and only if the individual using the expressor "positive integer" knew the unique factorization theorem or for some other reason believed that positive integers have this property.

The term "intensional" is used by Korzybski and Hayakawa in a way quite different from the present usage of "defining c-intension," though their usage does have some similarity to that of "incidental c-intension."

[10] One should not assume too hastily that because certain abstract entities do not exist in a spatio-temporal sense they are thereby eliminable in a theoretical analysis. The

There are two methods of investigating c-intensions. Perhaps the most obvious is to ask the subject what he means by a given expressor, thereby obtaining a *formulated c-intension* (a *definition* is a certain kind of formulated c-intension, namely, one which formulates a set of properties both *necessary and sufficient* for class membership; cf. Morris's "formulated significatum" [24]). However, many laboratory investigations as well as everyday observations have shown that formulated c-intensions may be very inadequate. Several investigators (9, 10, 11, 25, 28, 29) have shown this to be the case even in situations where the defining c-intensions can easily be formulated in the language available, which is not the case with certain everyday expressors, such as "dog." Exactly what is meant by "adequacy of a formulated c-intension" will be stated below. At any rate, the obvious inadequacy of formulated c-intensions necessitates the use of the second method, which is *inference from properties which c-denotata have in common,* at least those properties which were *easily perceivable* by the subject during the conditions under which the objects were made c-denotata. It should be emphasized that not all properties held in common by the c-denotata of a category are included in the c-intension of that category; only those which are perceived, believed, or conceived are so included. Further, some properties which the c-denotata do not actually have may be part of the c-intension. When an individual learns that the c-denotata have a property he did not previously know about, that property may become part of the incidental c-intension, though with time it may be absorbed into the defining c-intension.

With the preceding terms we can differentiate several kinds of content similarity between two categories. First, *c-extensional equivalence* [11] with respect to a set of objects *S*, which is simply the degree of extensional overlap (intersection of the two extensions), definable quantitatively. Second, *c-intensional equivalence,* or degree of similarity in c-intensions. In some cases this can be split up into defining c-intensional equivalence [12] and incidental c-intensional equivalence. These terms can also be used with reference to *concepts.* The c-extensional equivalence of two concepts with respect to a set *S* is the overlap [13] in the two respective partitions (sets of c-extensions). The c-intensional equivalence of two concepts is the similarity in the sets of c-intensions of the respective sets of categories.

It was stated earlier that a category may have more than one expressor. Given two expressors, "A" and "B," how do we decide whether they express the same category? There are several criteria to use:

1. Perfect c-extensional equivalence of "A" and "B" with respect to any set of objects.

2. Perfect c-intensional equivalence, in so far as it can be inferred from c-

reader interested in logic and mathematics is referred to Alonzo Church's paper, "The need for abstract entities in semantic analysis," *Proceedings of the American Academy of Arts and Sciences,* Vol. 80, No. 1, July, 1951, for discussion of a related point. For some of the circumlocutions which are necessary to avoid the use of certain abstractions in mathematics, see the paper by Goodman and Quine, "Steps toward a constructive nominalism," *J. symbolic Logic,* 1947, 12, 105–122.

[11] Cf. Carnap's "equivalence" (3), Johnson's "extensional agreement index" (12).

[12] Cf. Carnap's "L-equivalence," which is an either-or kind of predicate not involving similarity.

[13] The overlap of two partitions of a set can be precisely defined in several ways. For the purposes of this paper a definition is not necessary.

extensional equivalence and from formulated c-intensions.

3. A decrease in time required to decide whether to apply expressor "B" if the decision whether to apply expressor "A" has already been made, i.e., a decrease, compared with the time required to make the decision for similar objects not previously decided upon with respect to "A."

4. Indication from the subject, verbal or otherwise, that for objects for which the decision whether to apply "A" has already been made, the decision whether to apply "B" has also been made and explicitly indicated. In other words, that "B" "has the same meaning as 'A.'"

If all these criteria are met, good evidence has been obtained that "A" and "B" express the same category. In some cases criteria 1, 2, and 4 are sufficient, as with some expressors and some sets of objects decision time is so short anyway that no decrease can be observed.

Even if "A" and "B" are not expressors of the same category, they may be expressors of the same concept, i.e., "A" and "B" may be different yet both be categories of "C." It was asserted earlier that there are methods for determining this. We can now list the criteria for deciding whether "A" and "B" are *nonoverlapping categories* of "C":

1. Zero (or near zero) c-extensional equivalence.

2. A decrease in time required to decide whether X is a c-denotatum of "B" when it has already been decided whether X is a c-denotatum of "A."

3. An indication, verbal or otherwise, from the subject that if X has already been made a c-denotatum of "A" then the decision that X is *not* a c-denotatum of "B" has already been made and explicitly indicated.

4. An indication from the subject that he feels or believes that the judgment of whether X is a c-denotatum of "A" is the same kind of judgment as that of whether X is a c-denotatum of "B." (Don't depend on the subject, however, to explicate what he means by "same kind"!)

The fulfillment of these criteria constitutes evidence that "A" and "B" are expressors of *nonoverlapping categories* of "C." The problem of *overlapping* categories or of hierarchical structure is more complicated and no attempt will be made to list criteria here.

Most psychologists would probably agree that almost any word and the same word with "not," or the prefix "un" or "in" preceding it are expressors of the same concept. By using the above criteria we can verify this more or less obvious intuition, at least with human subjects. Certainly any expressors of the form "E" and "Not E" will have zero, or near zero, c-extensional equivalence. This is one way of stating the law of contradiction. It will in general take a subject less time to decide, for certain kinds of objects at least, whether to call an object "E" if he has already decided whether to call it "Not E." Further, if the expressor "E" has already been applied to an object, it will seem to the subject that the same kind of judgment is called for, that he has already decided not to apply the expressor "Not E," and that he has already explicitly indicated that he would not apply "Not E." Words called "opposites" also tend to satisfy these criteria.

Notice that there is no mention above of the following condition: if X is not a c-denotatum of "A" then X *will be* a c-denotatum of "B." This condition will be fulfilled in certain cases, but certainly not in general. Even expressors of the form "E" and "Not E" will not always fulfill this condition, e.g., merely because an individual will not apply the expressor "liberal" to a

given person does not mean that he will apply the expressor "not liberal," and this may hold even if he considers the liberality of the person and tries to make a decision. In other words, in pragmatics a law of excluded middle does not hold.

The foregoing discussion of content description makes it easier to discuss the *formal* properties of concepts. The following is a list of some of these properties:

1. *Structure.* By this is meant simply the number of categories and the relation between them. Possibly there exist concepts with only one category, e.g., concepts expressed by "entity," a word which is applied by some writers to any object, relation, etc., abstract or concrete. Most concepts, however, have at least two categories (though it may be that for many concepts only one category has any denotata), which is the simplest case of *nominal* or qualitative structure. Some nominal structures may be rather complex, as when one category includes two or more other categories. A fairly common hypothesis among psychologists is that, as the individual develops, his concepts have a more and more complex structure (cf. Welch and Welch and Long on hierarchical structure [34, 35]).

Some other types of structure are *ordinal, interval,* and *ratio.* It may seem surprising that these terms, developed in the theory of measurement (30), should turn up in this context. Actually, however, the psychology of concepts and the theory of measurement are intimately related, in fact, a really adequate theory of measurement requires an adequate psychology of concepts, whether one calls it that or not, a fact overlooked by some writers in the former field. In the case of interval and ratio operators, and sometimes with ordinal or nominal operators, the operation of looking or otherwise directing the sense organs and making a judgment is supplemented by certain prescribed manipulations of a more or less complex kind, but in all cases the end result is a category or value, and the problem of adequate description in terms of c-extensions, c-intensions, formulations, equivalences, etc., is just as much present in the case of interval and ratio operators (the problem of what the dimension means) as in the case of nominal operators.[14]

It seems apparent that there are important individual differences in operator structure. Frenkel-Brunswik (6, 7) has shown that some children think in dichotomies (weakness-strength, masculine-feminine, right-wrong) more than others, and she maintains that such dichotomous operators are generally characteristic of the ethnically prejudiced and the authoritarian personalities. Klein (13) maintains that a fundamental aspect of personality is revealed in individual differences in operator structure. Korzybski attributed many ills of the world to dichotomous operators ("two-valued, either-or orientations"), and some psychologists more or less agree with this kind of assertion.

2. *Decidability.* We call a concept *decidable* with respect to an argument X if X can be categorized, in a way satisfactory to the individual, by that concept acting as an operator, otherwise *undecidable.* We call X *c-decidable* or *c-undecidable.* The word "vague," when applied to concepts, sometimes means c-undecidability, though other meanings are low c-extensional equivalences and inadequacy of formulated intensions (see below). The inability to tolerate c-undecidability is probably related to intolerance of ambiguity as discussed by Frenkel-Brunswik (6, 7);

[14] Stevens (30) mentions only the most uninteresting case of nominal structure, that in which labels (numerals) are used for identification marks, without meaning (intension).

in her discussion, however, she stresses dichotomous (or other relatively undifferentiated) operators rather than extreme c-undecidability.

3. *Availability.* This is inferred from the conditions under which a concept will or will not operate. A measure of availability can be the amount of information necessary for operation, the speed of operation, etc. Availability is obviously related to the Höffding function.

In an experiment using incomplete figures similar to those of the Street Gestalt Completion Test (31), Verville and Cameron (32) found age and sex differences in the speed of operation of concepts (i.e., time between presentation of a stimulus and the perception of it as a certain kind of thing). Whether this result would hold with other stimulus materials and other conditions of operation has not, to the present writer's knowledge, been investigated, nor has the extent or significance of individual differences in this property been demonstrated. The speed of operation of a concept is analogous to the latency of a response, but this property has not been fully exploited either as a measure of how well a concept has been formed, analogously to the use of latency as a measure of habit strength, or as a diagnostic measure in personality investigations.

4. *Adequacy of formulation of the c-intension.* If an individual formulates in the language L the c-intension of his concept C, then the interpreter I *forms* a concept by *interpreting* the formulated intension. We can distinguish three kinds of adequacy:

a. C-extensional adequacy with exspect to a given set S. This is the c-extensional equivalence of the two concepts, C and the interpretation (in L) of the formulated c-intension. Smoke (28) found that some of his subjects' formulated intensions were not completely c-extensionally adequate, *even with respect to the set of objects used as arguments in the formation of the concepts.*

b. Defining c-intensional adequacy. This is the defining c-intensional equivalence of the two concepts.

c. Incidental c-intensional adequacy. This is the incidental c-intensional equivalence of the two concepts.

In scientific arguments over definitions considerable confusion results from failure to specify the kind and degree of adequacy expected or required. This point could with justification be considerably elaborated.

5. *Degree to which the c-intension is conscious.* To some psychologists this property will seem either unintelligible or reducible to property 4. The writer does not share this view; any close tie-up of the psychology of concepts with either psychoanalytic theory or with phenomenology presupposes an investigation of this property.

6. *Validity (with respect to a given set of objects).* The extent to which the c-denotata of the categories *actually satisfy* the respective c-intensions. If one categorizes, as in the Ames demonstration (20) a distorted room as a rectangular room, for this particular object in this situation his operator has very low validity. The validity of operators is one (but not the only) aspect of contact with reality.

7. *Domination.* The extent to which various behaviors and experiences of the individual are functions in a given situation of the result of the operation of the concept. We call such a concept a *dominator* in that situation. Three kinds of domination can be distinguished: *behavior, affect,* and *perception.* Refusing to shake hands with a Negro, feeling negative toward him, and seeing something wrong in everything he does are often examples of behavior,

affect, and perception dominators, respectively.[15]

We can give a partial description of a paranoid condition by saying that the individual is behavior, affect, and perception dominated by highly invalid operators.

In addition to the many problems in the field of concept formation which the foregoing distinctions suggest, we can state, using these distinctions, in fairly precise form a rather comprehensive set of theses about cognitive functioning as follows: Concepts which are describable in terms of formal properties and content[16] play a demonstrable role in the formation and organization of experience and behavior; the formal properties and the content of frequent operators are related to the personality structure of the individual; the content and formal properties of the operators frequently and characteristically used by members of a given culture are strongly related to the lexicon and grammar of the language used in the culture; potential dominators, i.e., concepts which if they operated would dominate, tend to be operators. If any of these theses is true, the psychology of concepts overlaps with many other areas of psychology and has important implications for many other disciplines and applied fields.

[15] It is well recognized that expressors of dominators are sometimes re-defined in an attempt to change the extensions and/or intensions and yet keep the domination. C. S. Stevenson, in his book *Ethics and Language*, calls such a definition a "persuasive definition."

[16] There is a school of thought which holds that concepts, if they are to be talked about at all, must be defined and described in terms of *behaviors* rather than in terms of formal properties and content as described in this paper. It is obvious that behavioral *indices* must be used (such as sorting, discriminatory responses, labeling, autonomic responses, etc.) in the determination of c-extensions and c-intensions; however, the present author believes that any attempt to *define* and *describe* concepts in terms of behaviors will turn out to be about as hopeless and fruitless as was the older attempt to describe concepts in terms of images. Behavior is the result of so many variables that it is difficult to see why anyone regards behavioral *definitions* of any one variable (of any appreciable complexity) as other than a hopelessly complicated task. It should be noted that *descriptions existing in the present literature do describe concepts in terms of c-extensions and c-intensions*, though of course they use behavioral indices in obtaining these descriptions. For a statement of a point of view held by some structural linguists and some psychologists and rather divergent from that presented in this paper, see Bloomfield, L., "Language or ideas?" *Language*, 1936, **12**, 89–95.

A more serious objection to the present discussion runs as follows: the author seems to endorse the ancient view that concepts are entities which are carried around by the individual and applied to objects from time to time without taking into account the inner and outer situational contexts, which introduce many changes in these so-called entities. The author readily admits the force of this objection and regards the present treatment as a first approximation only, though one that can lead to important results.

REFERENCES

1. ALLPORT, G., & POSTMAN, L. *The psychology of rumor*. New York: Henry Holt, 1947.
2. BARTLETT, F. C. *Remembering*. Cambridge: Cambridge Univer. Press, 1932.
3. CARNAP, R. *Meaning and necessity*. Chicago: Univer. of Chicago, 1947.
4. DUNCKER, K. On problem solving. *Psychol. Monogr.*, 1945, **58**, No. 5 (Whole No. 270).
5. FREGE, G. Sense and reference. *Philos. Rev.*, 1948, **57**, 209–230.
6. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
7. FRENKEL-BRUNSWIK, ELSE. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 356–419.
8. FRIES, C. C. *The structure of English*. New York: Harcourt, Brace, 1952.
9. HEIDBREDER, EDNA. The attainment of concepts: I. Terminology and methodology. II. The problem. *J. gen. Psychol.*, 1946, **35**, 173–189, 191–223.

10. HEIDBREDER, EDNA. An experimental study of thinking. *Arch. Psychol.*, 1924, No. 73.
11. HEIDBREDER, EDNA. Toward a dynamic psychology of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.
12. JOHNSON, W. *People in quandaries.* New York: Harper, 1946.
13. KLEIN, G. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 328–355.
14. KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.
15. KOHLER, W. *Gestalt psychology.* New York: Liveright, 1947.
16. KOHLER, W. *Dynamics in psychology.* New York: Liveright, 1940.
17. KORZYBSKI, A. The role of language in the perceptual process. In R. Blake & G. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 170–205.
18. KORZYBSKI, A. *Science and sanity.* (3rd Ed.) Lakeville, Conn.: Int. Non-Aristotelian Library Pub. Co., 1948.
19. KRECH, D., & CRUTCHFIELD, R. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
20. LAWRENCE, M. *Studies in human behavior.* Princeton: Princeton Univer. Press, 1949.
21. LEEPER, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 730–757.
22. LEWIS, C. I. *An analysis of knowledge and valuation.* La Salle, Ill.: Open Court, 1946.
23. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
24. MORRIS, C. *Signs, language, and behavior.* New York: Prentice-Hall, 1946.
25. REES, H. J., & ISRAEL, H. E. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**, No. 6 (Whole No. 210).
26. SCHACHTEL, E. On memory and childhood amnesia. *Psychiatry*, 1947, **10**, 1–26.
27. SHERIF, M., & CANTRIL, H. *The psychology of ego-involvements.* New York: Wiley, 1947.
28. SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**, No. 4 (Whole No. 191).
29. SNYGG, D. The relative difficulty of mechanically equivalent tasks. I. Human learning. *J. genet. Psychol.*, 1935, **47**, 299–320.
30. STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1–49.
31. STREET, R. F. A gestalt completion test: a study of a cross section of intellect. New York: Bur. Publ., Teachers Coll., 1931.
32. VERVILLE, E., & CAMERON, N. Age and sex differences in the perception of incomplete pictures by adults. *J. genet. Psychol.*, 1946, **68**, 149–157.
33. VINACKE, W. E. The investigation of concept formation. *Psychol. Bull.*, 1951, **48**, 1–31.
34. WELCH, L. A preliminary investigation of some aspects of hierarchical development of concepts. *J. genet. Psychol.*, 1940, **56**, 359–378.
35. WELCH, L., & LONG, L. The higher structural phases of concept formation of children. *J. Psychol.*, 1940, **9**, 59–95.
36. WERNER, H. *Comparative psychology of mental development.* (Rev. Ed.) Chicago: Follett, 1948.
37. WHORF, B. An American Indian model of the universe. *Int. J. Amer. Linguistics*, 1950, **16**, 67–72.
38. WHORF, B. *Four articles on metalinguistics.* Foreign Service Institute, Dept. of State, 1949.

[MS. received June 9, 1952]

Psychological Review
Vol. 60, No. 4, 1953

EXPECTANCIES AND HULLIAN THEORY

RICHARD A. BEHAN

Michigan State College

The present paper is written to show that the Hullian theory of behavior is capable of including an expectancy concept which may be likened to that of Tolman (5). It is proposed to derive this concept of expectancy as a theorem to Hull's 1943 postulate set (3). Then, in the language that Hull has preferred, the notion of expectancy will be a corollary. It will be necessary to explain first what is meant by the word *expectancy* in the present paper, and how the notion will be conceptualized. Then the assumptions which are necessary will be set forth, followed by the deduction.

The starting point for the present notion will be the *sign-gestalt-expectation* of Tolman (5). We quote Tolman: "In the case of a sign-gestalt-expectation the organism is ready either as a result of perception (q.v.), mnemonization (q.v.), or inference (q.v.) to have some sort of positive or negative commerce with an immediately presented object (i.e., the sign-object) as the means-end-route (i.e., the signified means-end-relation) to get to or from such and such a further object (i.e., the signified-object)" (5, p. 453). It is thus seen that the sign-gestalt-expectancy is a relation between signs such that the first occurring sign elicits an approach to or an avoidance of the later occurring sign. Special emphasis is placed on the wholeness or the unity of this trinity, and the fact that it directs behavior.

If the above quotation were to be translated into the language of Hullian theory (it is hoped, without doing violence either to Hullian concepts, or to Tolman's terminology), it would read like the following: The goal object evokes a goal response. On successive trials stimuli in the immediate environment come to signal the occurrence of the goal object. Gradually, as the number of trials increases, stimuli which are more distant from the goal object acquire this property of signaling the occurrence of the goal object. Eventually the entire experimental apparatus will possess directive value for the organism. The directive properties which these stimuli acquire result from their relation with the goal object (incentive stimulus). It is this relationship between the incentive stimulus and other stimuli in the experimental situation that constitutes the expectancy.

The expectancy will be conceived as a power of a relation between relations which hold between stimuli and responses, i.e., the expectancy is seen as a relationship which holds between two relations which take stimuli and responses as arguments.[1] First, the relation *evokes,* which holds between a stimulus complex and a response complex, will be considered to be irreflexive and asymmetrical. Second, the relation *leads to,* which holds between response complexes and stimulus complexes, will be considered to be irreflexive and asymmetrical also. The expectancy will then be understood as a power (*po*) of the relation *relative product* holding between the relation *evokes* and the relation *leads to.*

The notion of the relative product is introduced by Whitehead and Russell in their *Principia Mathematica* in Chapter 34 of Volume I (6). The relative

[1] The arguments of a relation are the terms that are related by that relation.

product is a relation between two relations when there is some term to which the members of the domain of the first relation are related by the first relation, and to which the members of the converse domain of the second relation are related by the second relation. That is to say, if we let "R" and "S" denote relations, and "(R/S)" denote their relative product, then the relative product may be defined as

$$(R/S) = {}_{Df}\hat{x}\hat{z}\{(\exists y)\cdot xRy\cdot ySz\}$$

The power of a relation is the relation between two terms when one is a finite number of relation steps removed from the other. We quote Woodger who says of the relation: "R_{po} is the relation of being related by some power of R. Thus we have $xR_{po}y$ when y can be reached from x in some finite number of R-steps, i.e., when

$$xRv\cdot vRw\cdot \ \ldots\ \cdot zRy$$
$$\text{or } x(R\cup R^2\cup R^3\cup \ldots)y"$$

(7, p. 39). R_{po} is defined as follows:

$$R_{po} = {}_{Df}\hat{x}\hat{y}\{(u)\cdot (\breve{R}``u \subset u . \overleftarrow{R}`x \subset u) \supset y \in u .\}$$

R_{po} will be the relation of x to y if it is the case that, whatever class u may be, if $\breve{R}``u$ are included in u and if the $\overleftarrow{R}`x$ are included in u, then $y \in u$."

We are now ready to start the derivation. It is necessary first to find in Hullian theory the relation *evokes* and the relation *leads to*. The first of these relations is easily found in postulate 11 of *Principles of Behavior* (3). Postulate 11 asserts "The momentary effective reaction potential ($_S\vec{E}_R$) must exceed the reaction threshold ($_SL_R$) before a stimulus (S) will evoke a response (R)." This postulate provides the relation *evokes*, although the reader will note below that we have taken only the word statement of the postulate. The symbolic statement given in *Principles of Behavior* is defective and we prefer the sentence of assumption 5 to the one listed by Hull in his book. This postulate 11 is now contained in the B part of the 14th postulate in the new postulate set (4).

The notion of a response leading to a stimulus complex is not so easily found. This notion does not appear explicitly in Hull's theory, although it is introduced informally (and incidentally is assumed by many of the students of Hull—and students of students of Hull). For example, the diagram on page 96 of *Principles of Behavior* implies that response may lead to stimulation. Also in some of his earlier work, see especially (2), Hull assumes that response will lead to stimulation. At this point, then, two courses are open: (*a*) One may proceed on convention and simply assume that the Hullian theory contains the notion of a response's leading to stimulation. Or, (*b*) one may stay within the formal theory and derive this notion by the method of substitution. The latter alternative has been chosen in this case for the reason that it is desirable to remain within the theory in the present derivation, and, despite the wide acceptance of the convention mentioned in alternative (*a*), the acceptance of this convention goes beyond the formal theory. It was felt that it was better not to go beyond the theory in cases where it is desirable to show that the theory (the formal theory) contains the notions with which one wishes to work.

Corollary XII in the chapter on *Inhibition and Effective Reaction Potential* asserts: "Whenever conditioned reactions are evoked, whether reinforced or not, reactive inhibition (I_R) is generated" (p. 290). We shall substitute in this corollary the notion of a stimulus complex for the notion of reactive inhibition, obtaining the following lemma:

Lemma I. Whenever conditioned reactions are evoked, whether reinforced or not, stimulus complexes are generated. We may now define the notion of a response leading to stimulation as equivalent to lemma I.

Following is a list of the assumptions which are necessary for the derivation which follows:

1. The logical type of the symbol which denotes *reactive inhibition* is the same as the logical type of the symbol which denotes *stimulus complex*. This assumption is necessary for the justification of the substitution through which we arrived at lemma I from corollary XII.

2. Assume an experimental situation such that the stimulus complexes may occur in serial order, and assume also a succession of trials in this experimental situation. The serial order may be temporal only, or it may be temporal and spatial.

3. Assume that the effective reaction potential of responses with which we wish to deal is in excess of the reaction threshold in each case.

4. Assume the notions of reinforcement and secondary reinforcement.

5. Assume postulate 11.

6. Assume corollary XII.

7. Assume that if the i^{th} stimulus evokes a response, then this response is the i^{th} response.

8. Assume that if the i^{th} response leads to a stimulus complex, then this is the $i+1$st stimulus complex.

Assumptions (7) and (8) specify properties of the experimental situation.

We may now proceed directly to the derivation. Some of the assumptions will be listed below in symbolic language, although not all will enter directly into the derivation. We let "j," "k," "l," and so forth be variables for number signs,

"s_j" denote the jth occurring stimulus complex in the experimental situation,
"r_j" denote the jth occurring response complex in the experimental situation,
"s_i" denote the incentive stimulus, and the i^{th} occurring stimulus complex,
"s_s" denote the stimulus complex at the beginning of the experimental situation, and the s^{th} occurring stimulus complex,
"$s_j \phi r_j$" denote "the stimulus complex s_j evokes the response r_j,"
"$r_j \psi s_k$" denote "the response r_j leads to the stimulus complex s_k."

The symbolic statements of the necessary assumptions follow:

Assumption 3.	$(j,k) s_j E r_k > s_j L r_k$	
Assumption 5.	$s_j \dot{\hat{E}} r_j > s_j L r_j . \supset . s_j \phi r_j$	
Assumption 6.	$(x):(\exists y).x \epsilon R.y \epsilon I_R. \supset .Ex \supset Gy$	
Assumption 7.	$(j,k).s_j \phi r_k \supset j = k$	
Assumption 8.	$(j,k).r_j \psi s_k \supset k = j+1$	
Derivation:		
Postulate 11	$s_j \dot{\hat{E}} r_j > s_j L r_j . \supset . s_j \phi r_j$	(1)
Assumption 3. (Spec.)	$s_j \dot{\hat{E}} r_j > s_j L r_j$	(2)
(1).(2).Modus Ponens	$s_j \phi r_j$	(3)
Corollary 12	$(x):(\exists y).x \epsilon R.y \epsilon I_R. \supset .Ex \supset Gy$	(4)
(4). $[S/I_R, z/y, \hat{u}(Gu)/G]$	$(x):(\exists z).x \epsilon R.z \epsilon S. \supset .Ex \supset \hat{u}(Gu)z$	(5)
(5). (Concretion)	$(x):(\exists z).x \epsilon R.z \epsilon S. \supset .Ex \supset Gz$	(6)

Define	$r_j\psi s_i = {}_{Df}(x):(\exists z).x\epsilon R_j.z\epsilon S_i.\supset.Ex\supset Gz$	(7)
(6).(7)	$r_j\psi s_i$	(8)
(3).(8)	$s_j\phi r_j.r_j\psi s_i$	(9)
(9).[$j/i.k/j$]	$s_k\phi r_k.r_k\psi s_j$	(10)
(9).[$1/j.k/i$]	$s_1\phi r_1.r_1\psi s_k$	(11)
	.	
	.	
	.	
(9).[$s/j.1/i$]	$s_s\phi r_s.r_s\psi s_1$	(12)
(12).(11).(10).(9).	$s_s\phi r_s.r_s\psi s_1.\ldots.s_1\phi r_1.r_1\psi s_k.$	
(Assumption 2)	$s_k\phi r_k.r_k\psi s_j.s_j\phi r_j.r_j\psi s_i$	(13)
Define[2]	$(\phi\|\psi) = {}_{Df}\hat{s}_m\hat{s}_n\{(\exists r_m):s_m\phi r_m.r_m\psi s_n\}$	(14)
(13).(14)	$s_s(\phi\|\psi)s_1.\ldots.s_1(\phi\|\psi)s_k.$	
	$s_k(\phi\|\psi)s_j.s_j(\phi)\|\psi)s_i$	(15)
(15).	$s_s(\phi\|\psi)_{po}s_i.$	(16)

(16) is a relation between the relation *evokes* and the relation *leads to*, i.e., (16), which is the relative product of these two relations, is the expectancy. It will be noted that (15) is just as good as (16), indeed, it is a good deal more expressive. One may think of the process of expectancy development as a process which results in bringing more stimulus complexes into a particular relation with the goal stimulus (incentive stimulus). The expectancy, as it is developed in the present case, is a relation mediated by response, taking stimulus complexes as arguments.

In the quotation from Tolman in the second paragraph of this article the following phrase appears: "In the case of a sign-gestalt-expectancy the organism is ready either as a result of perception (q.v.), mnemonization (q.v.), or inference (q.v.) to have some sort of commerce . . . ." Each of these three terms (perception, mnemonization, and inference) designates a mode of sign-gestalt-expectation. The present writer takes the liberty of assuming that in any actual situation all three modes are active. It is assumed further that Tolman is here referring to a completed expectancy, such as might be similar to our (16). According to lemma I, it is reasonable to assume that during the development of the expectancy the mode of inference plays a very important part.

At this point, a few remarks about lemma I might well be in order. It is obvious that this lemma depends upon corollary 12. If it should ever become necessary to drop corollary 12 from the theory, then lemma I would be eliminated. In such a case, if it were desirable to keep the present notion of expectancy, it would be necessary to make the statement that responses lead to stimulation a part of the formal theory. Such a statement in the form of a postulate would be of great use to a 'theory like Hull's. For example, such an assumption would allow the

[2] This is the definition of relative product which appears in *Principia Mathematica* (6). The symbolization has been modified to suit the purposes of the present usage.

derivation as a corollary of the postulate of behavioral oscillation (postulate 10) (3).

The relationship between the notion of expectancy, as here presented, and the rest of Hullian theory must now be considered. Every useful construct must have an operational definition and must be related to other constructs in the theory of which it is a part. The beauty of such a derivation as that given above is that it necessitates no new additions to the list of operational definitions in a theory. The derivation provides the necessary reduction chain to reduce the new concept (1). The reduction basis for expectancy in Hullian theory will be the same as that of ${}_{S}\hat{E}_{R}$. The derivation also provides the necessary connections with other constructs in the theory. Because the construct expectancy springs from ${}_{S}\hat{E}_{R}$, it will automatically become connected with all of the constructs in Hull's theory which contribute to ${}_{S}\hat{E}_{R}$.

The present paper will close with a remark about the importance of the possibility of such a derivation as this paper contains for one of the traditional areas of friction in psychology. Here reference is made to the molecular vs. molar controversy. These two terms refer to relative differences between levels of analysis and not to any real fixed differences between approaches to the subject matter of psychology. Those individuals who begin with a relatively molar analysis of the phenomena of psychology end up with a relatively molar analysis of the phenomena they study. Those individuals who begin with a relatively molecular analysis of the phenomena of psychology may, if they utilize the proper techniques, also have a relatively molar analysis of the phenomena they study. The molecularites can have the molar cake and eat it too.

REFERENCES

1. CARNAP, R. Testability and meaning. *Phil. Sci.*, 1936, **3**, 420–471; 1937, **4**, 1–40.
2. HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **37**, 511–525.
3. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. HULL, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
5. TOLMAN, E. C. *Purposive behavior in animals and men.* Berkeley: Univer. of Calif. Press, 1932.
6. WHITEHEAD, A. N., & RUSSELL, B. *Principia mathematica.* Vol. 1. Cambridge: Cambridge Univer. Press, 1927.
7. WOODGER, J. H. *The axiomatic method in biology.* Cambridge: Cambridge Univer. Press, 1937.

[MS. received July 12, 1952]

Psychological Review
Vol. 60, No. 4, 1953

STIMULUS SATIATION: AN EXPLANATION OF SPONTANEOUS ALTERNATION AND RELATED PHENOMENA [1]

MURRAY GLANZER

Brooklyn College

In 1925 Tolman (20, p. 290) described a peculiar regularity in the behavior of rats that was later labeled spontaneous alternation.

> A simple T maze was used, and it was arranged so that the animal could get back to the food box in identical fashion whether he chose the left or the right angle of the T. Either route met with success. . . . [There was] a very pronounced tendency toward continuous and regular alternation —left, right, left, right, or right, left, right, left. . . . It appeared, in short, that even where either side was equally "satisfactory" there was in our rats a positive tendency left over toward variation of response . . . a positive tendency in and of itself.

Ten years after Tolman's report, spontaneous alternation received intensive study by Dennis and his coworkers. Although these early investigators did not incorporate their data into a formal theory, Dennis (5) did imply that spontaneous alternation was in some way determined by the stimuli the subject faced. For example, the title of one article was "Spontaneous Alternation in Rats as an Indicator of the Persistence of Stimulus Effects," and, in discussion of his results, he wrote about "a tendency to avoid a specific pathway which has recently been traversed" (5, p. 310). This formulation suggests a theory of spontaneous alternation in terms of the stimulus. The rat alternates from one alley (a set of stimuli) to the other (a different set of stimuli), going to Alley B because it has already experienced Alley A. The theory proposed in this paper develops and systematizes this basic approach.

In direct antithesis to a stimulus-oriented theory is one which interprets spontaneous alternation in terms of the response. In a response theory, the animal is viewed as alternating responses instead of alleys, i.e., making a right turn because it has already made a left turn. Many of the more recent investigations examined below have stemmed from this theoretical position.

A Critique of Response Theories

Explanation of spontaneous alternation in terms of the response has been attempted by several Hull-influenced investigators: Heathers (9) using the concept of performance decrement; and Solomon (19) and Zeaman and House (23) using the concept of reactive inhibition. According to both concepts, the occurrence of a response reduces temporarily the probability of its recurrence. The concepts were employed to explain both the fact of spontaneous alternation and the fact, predicted and demonstrated by Heathers (9), that spontaneous alternation disappears as the interval between trials increases.

The concept of reactive inhibition, in addition, assigns an explicit role to the number of times the response is made and the amount of work required by the response. Thus, Zeaman and House (23) predicted and found that giving animals forced trials to one of two

[1] This paper is based upon a Ph.D. dissertation submitted in February 1952 to the department of psychology at the University of Michigan. The writer is indebted to the members of the committee and to Dr. Edward L. Walker, chairman of the committee, who gave invaluable aid in the preparation and editing of this paper.

alleys increased the tendency to alternate to the other; Solomon (19) predicted and found some positive but not conclusive evidence that increased work resulted in increased spontaneous alternation.

Although these experiments seem to give considerable support to an explanation in terms of the response, a number of facts remain that resist explanation by this type of theory. Typical of these are the following:

1. Dennis (5) found that rats did not show spontaneous alternation from unit to unit of a multiple-unit maze. That is, they did not go right-left-right-left within a single trial on a four-unit maze, but showed instead spontaneous alternation in their successive choices at a given choice point. They might go right-right-right-left on the first trial and left-left-left-right on the second trial. This finding is in direct contradiction to the prediction which should follow from a theory in terms of the response.

2. Jackson (11) found that the difference between the two possible responses at the choice point did *not* affect the amount of spontaneous alternation. A Y maze with a small angle of separation, requiring a choice between two practically identical responses, elicited as much spontaneous alternation as a Y make with a wide angle of separation. If the response were the key to spontaneous alternation, then differentiating the response should result in more alternation.

3. Zeaman and House (23), using a procedure of ten forced trials followed by a free trial, and Walker,[2] using simply two free trials, found that rats showed spontaneous alternation with intervals greater than an hour between trials. These findings disagree with expectations on the basis of reactive inhibition which is ordinarily assumed to dissipate within a few minutes.[3]

4. Wingfield (21) found that human subjects showed more spontaneous alternation in choice of lights that differed in color than of lights that were the same. Something in addition to consideration of the subject's responses seems to be demanded by this finding.

5. Dennis (5) varied the amount of work for rats running a maze by increasing the length of the final section from zero to 38 ft. but found no change in the amount of spontaneous alternation. It would seem reasonable to expect that the reactive inhibition from the added run of 38 ft. should dwarf the effects of reactive inhibition from the momentary act of turning right or left.

These difficulties suggest that it might be fruitful to turn to the construction of an explanation in terms of the stimulus. As already pointed out, Dennis (5) anticipated this type of explanation. Recently Montgomery (16, 17, 18), on the basis of empirical findings of his own, has questioned the reactive-inhibition explanation of spontaneous alternation and proposes the concept of an exploratory tendency which is reduced by exposure to a pathway of the maze. Berlyne (2) has presented a similar construct, a curiosity drive, designed to explain curiosity in rats. (Working toward a very different goal, that of incorporating perception and attention as "response" in Hullian theory, Berlyne (3) has also sketched the outlines of a theory that displays some points of similarity to the formulation presented below.)

[2] Walker, E. L. Unpublished data.

[3] Reactive inhibition has to be assumed to disappear quickly in order that it meet its other theoretical commitments. (See Zeaman and House [23].)

A THEORY IN TERMS OF THE STIMULUS

We shall assume that, with continued exposure to an *environment*—to the same *stimuli*—the organism becomes less active in that environment. With an eye to several experiments (**5, 9, 19, 23**), we assign a specific role to time in the disappearance of this boredom-like effect. Since, moreover, the explanation is in terms of the stimulus, it would be expected that "boredom" created in one situation would carry over to other similar environments and that the more similar two environments are, the greater the carryover.

Using these ideas to form the construct of stimulus satiation,[4] we shall follow Hull's procedure (**10**) and present the basis of the theory in the form of a postulate.

This new postulate may not only enable us to circumvent the difficulties of the reactive-inhibition explanation but will also yield novel predictions. Furthermore, it may serve to interrelate and explain empirical results from a number of apparently unrelated areas.

The postulate reads as follows:

Each moment an organism perceives a stimulus-object or stimulus-objects, A, there develops a quantity of stimulus satiation to A.

i. *The same amount of stimulus satiation develops in each successive moment. The total amount developed is, therefore, an increasing linear function of time.*

ii. *There is loss of part of each quantity of stimulus satiation in each successive moment. The amount of stimulus satiation remaining from each quantity is a decreasing negative exponential function of time.*

iii. *Stimulus satiation developed to A will be generalized to other stimulus-objects B. The amount of generalized stimulus satiation is an inverse function of the discriminability of A and B.*

iv. *The various quantities of stimulus satiation combine additively.*

v. *Stimulus satiation reduces the organism's tendency to make any response to A.*

Definition 1. Generalization. When a change in behavior toward A results in a similar change in behavior toward B, all other things being equal, generalization has taken place from A to B.

Definition 2. Discriminability. The ease with which a subject can be induced to make a different response to A than to B, all other things being equal, is called the discriminability of A and B.

Two corollaries follow from the postulate:

Corollary I. As long as an organism remains perceiving A, the amount of stimulus satiation it has to A at a given moment (that is, the total amount developed [i] minus the total amount lost [ii]), is an increasing negative exponential function of time.

Corollary II. When an organism stops perceiving A, the amount of stimulus satiation it has to A at a given moment is a decreasing negative exponential function of time.

These corollaries are presented in Table 1 and Fig. 1 with arbitrary values assigned to the functions.

Given the characteristics of various situations, the postulate yields a large number of deductions. A number of these for some basic types of situations are formulated and discussed below. For each deduction discussed, we will indicate whether or not it is crucial for the two theories, e.g., one in terms of the stimulus as against one in terms of the response, as well as state the nature of any existing evidence.

DEDUCTIONS FOR APPROACH RESPONSES IN A TWO-ALTERNATIVE SITUATION

Spontaneous Alternation

Let us start with a very simple situation. This situation is one in which *S* has evenly spaced, consecutive trials in

[4] This term is suggested by Karsten's work (**12**).

TABLE 1

COROLLARIES I AND II ILLUSTRATED BY OBTAINING VALUES FOR STIMULUS SATIATION, $I_n = 100\ e^{-(n-1)}$ (WHERE n IS THE NUMBER OF MOMENTS SINCE THE APPEARANCE OF THE STIMULUS)*

Stimulus Satiation Quantity	Moments							
	Stimulus Present				Stimulus Absent			
	1	2	3	4	5	6	7	8
I_1	100.0	36.8	13.5	5.0	1.8	.7	.3	.1
I_2		100.0	36.8	13.5	5.0	1.8	.7	.3
I_3			100.0	36.8	13.5	5.0	1.8	.7
I_4				100.0	36.8	13.5	5.0	1.8
ΣI_n	100.0	136.8	150.3	155.3	57.1	21.0	7.8	2.9

* Each row gives the history of a single quantity of stimulus satiation as it decreases with time. Each column gives the total amounts of stimulus satiation present during a single moment.

an apparatus that contains two distinct areas, such as a single-unit T or Y maze. A trial starts with *S*'s entrance into the apparatus and ends with its exit from the apparatus.

In order to complete a trial *S* has to enter one of the two areas or alleys. These areas should be initially more or less equally attractive to *S*. During the course of the experiment, furthermore, one area should not be made more or less attractive than the other by introducing reward or punishment. A situation that has these characteristics will be called a simple two-alternative situation.

The alternatives correspond to stimulus-objects A and B of the postulate. The *S* approaches and perceives one of the two on each trial. It will be assumed that *S*, upon entering an alternative, "perceives" it.

The *S*, perceiving one alternative, builds up stimulus satiation to it that

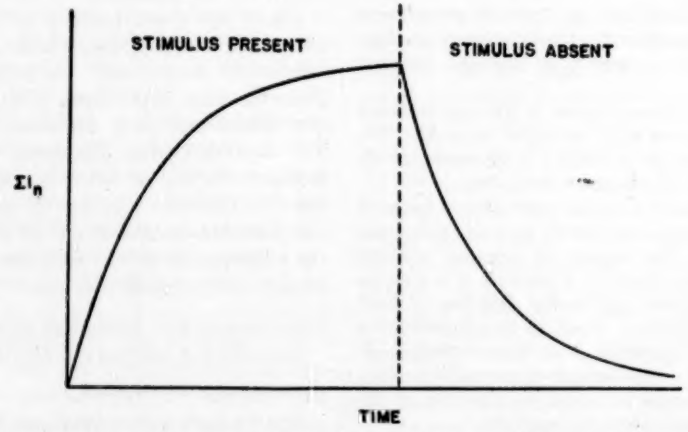

FIG. 1. Total amount of stimulus satiation, ΣI_n, as a function of time and the presence or absence of the stimulus-object. The totals of the columns of Table 1 are represented by this figure.

reduces the tendency to respond to that alternative. The response considered here is the response of approach, since *E* usually records and requires only that response for completion of a trial.

Deduction 1. In the simple two-alternative situation, the subject will alternate choice of arms or alleys in successive trials.

Verified; not crucial.

This is the simple fact of spontaneous alternation which has been repeatedly demonstrated (4, 5, 9, 11, 19, 20, 22, 23).

Deduction 2. In the simple two-alternative situation, if there is an exchange between trials of stimulus-objects so that the cues that were on one side of the *S* now are on the other, and vice versa, *S* will alternate approach to stimulus-objects in successive trials. This will mean a *repetition* of responses rather than an alternation of responses (turns).

Verified; crucial.

This is a key deduction since it contrasts the stimulus satiation and reactive-inhibition explanation most sharply. Stimulus satiation theory predicts a reduction in alternation of response, whereas a theory in terms of the response predicts no decrease.

The author (8) ran rats in a cross-shaped maze with two starting boxes (north and south), making it possible for the *S*s to start from one starting box for one trial and from the opposite starting box for the next. When this was done, the position of the alleys relative to *S* was reversed: the alley that had been on its right during Trial 1 was on its left for Trial 2. By thus shifting stimuli, it could be determined whether the animals alternated responses as required by the reactive-inhibition explanation or alleys as required by the stimulus satiation explanation.

Twenty-six rats were given two immediately consecutive trials a day for eight days. On four of the days both trials were from the same starting box; on the other four days the trials were from opposite starting boxes. As predicted above, the animals tended to repeat (i.e., go right-right or left-left) rather than alternate their responses on the days that the cues were shifted. This indicates clearly that they were alternating alleys (stimuli) rather than turns (responses). A similar experiment has been described by Montgomery (18) who reports results consistent with these findings.

Deduction 3. In the simple two-alternative situation if an extended series of consecutive trials is given, the tendency toward spontaneous alternation will be at a maximum for the first pair of trials, and will decrease to a minimum in later trials. This decrease will be called cumulation-effect.

Tested but not clearly verified; not crucial.

In a series of consecutive trials in a two-alternative situation, if the first trial was to alternative A, then the second trial would be to alternative B (Deduction 1). There would be two stimulus satiation curves then: one for alternative A starting from the first trial, and continuing its rise through later trials to alternative A; the other for alternative B, starting from the second trial. Since, according to Corollary I, curves of stimulus satiation are negatively accelerated, the two curves would approach each other in the later trials. The difference in amount of stimulus satiation to the two alternatives would therefore decrease as illustrated in Fig. 2, leading to the predicted decrease in spontaneous alternation.

Although results from Wingfield and Dennis (22), Heathers (9) and the author (7) show the general trend toward decrease in spontaneous alternation from first to last trial in a series, the data also show reversals and considerable irregularity.

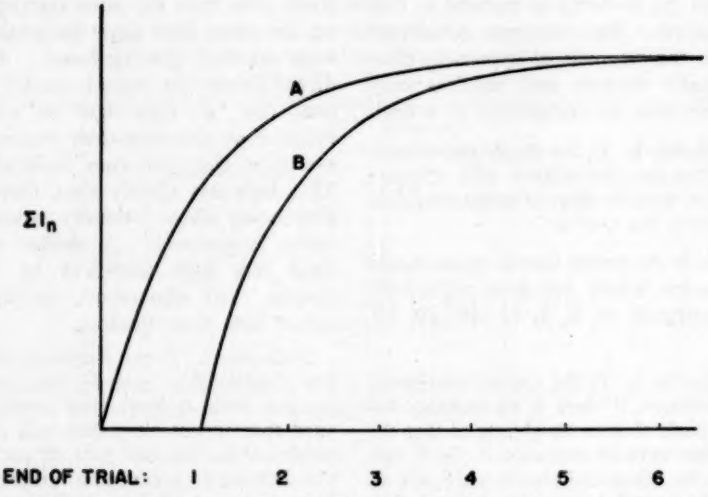

FIG. 2. Total amount of stimulus satiation, ΣI_n, to two alternatives, *A* and *B*, in successive alternating trials. The curves, approaching their limits, cause a decrease in the difference (the distance between the curves) in the amount of stimulus satiation to the two alternatives. All odd-numbered trials are to *A*, even-numbered to *B*. Decreases occurring during intertrial intervals have not been drawn.

The inconclusive nature of these data may result from the very procedure indicated by Deduction 3. In this procedure the animal is free to choose either alternative each trial in the series. The *E* therefore soon loses control of the basic variable, the number and order of particular choices that precede a given trial.

In order to make an adequate test of the prediction, it may be necessary either to use a group of animals large enough to cancel individual differences in the pattern of choices or to change the procedure in order to control the pattern of choices throughout. Clearer results would be expected, then, for the following deduction.

Deduction 4. In the simple two-alternative situation, if a number of forced consecutive alternation trials is given, as the number of trials increases, the tendency toward spontaneous alternation decreases.

Not tested; not crucial.

Now we turn to consideration of the role of generalization. On the basis of part iii of the postulate, the prediction could be made that the more discriminable the alternatives, the greater the amount of spontaneous alternation. Discriminability can be varied by several methods. One is to eliminate or add cues that differentiate the two alternatives.

Deduction 5. In the simple two-alternative situation, as stimuli differentiating the alternatives are eliminated, spontaneous alternation will decrease.

Tested but not verified; crucial.

A study on this by the author (7) comparing spontaneous alternation in Y mazes with many and few differentiating cues yielded negative results. In that experiment, however, there may have been inadequate control of the differentiating cues, since only intramaze cues were varied.

Since discriminability is a function of the organism as well as the environment, the following deductions can also be derived.

Deduction 6. In the simple two-alternative situation, if the ability of the organism to discriminate is reduced, spontaneous alternation will decrease.

Of the conditions of the organism that may affect its ability to discriminate, two may be singled out on the basis of empirical studies. One is sense-organ damage and the other is brain-injury (15). If these conditions may be considered to reduce discriminability, then the following more specific deductions may be made.[5]

Deduction 6*a*. In the simple two-alternative situation, the greater the extent of *S*'s sense-organ damage, the less the amount of spontaneous alternation.

Not tested; crucial.

Deduction 6*b*. In the simple two-alternative situation, the greater the extent of *S*'s brain damage, the less the amount of spontaneous alternation.

Verified; crucial.

Certain of Krechevsky's (13, 14) results support this deduction. He found that brain-operated rats perseverated in choice of one of two paths much more than did normal rats. His situation described below was somewhat complicated but the basic two-alternative arrangement was employed.

Now we turn to the effects of time on spontaneous alternation.

Deduction 7. In the simple two-alternative situation, as the interval between trials increases, the amount of spontaneous alternation decreases.

Verified; not crucial.

This deduction follows simply from part ii of the postulate. It has been verified by Heathers (9) who found that, as the time interval between trials increased from 15 sec. to 15 min., spontaneous alternation decreased from 86 to 50 per cent (chance). It is also verified by Zeaman and House (23), who found a decrease over a period of 24 hr.

Deduction 8. In the simple two-alternative situation, the more highly differentiated the alternatives, the longer the time interval between trials necessary for the disappearance of spontaneous alternation.

Not tested; crucial.

This may, in part, explain the widely discrepant data that have been obtained concerning the effects of delay on spontaneous alternation. For example, Walker,[6] using a highly differentiated Y maze, found spontaneous alternation with intervals longer than an hour. Montgomery (16), who took special care to keep his alternatives alike, found that spontaneous alternation disappeared with intervals as short as 90 sec.

Now we turn to an implication of Corollary I, that the longer the time the organism perceives a stimulus-object A, the greater the amount of stimulus satiation to A.

Deduction 9. In the simple two-alternative situation, if *S* is detained in one alternative, the amount of spontaneous alternation to the other alternative on the following trial will be greater, the longer the time of detention.

Not tested; crucial.

The apparent contradiction between Deduction 9 and Deduction 7 may be resolved in the following way. In Deduction 7, the concern is with "interval between trials," an interval spent *outside* the alternative chosen, whereas in Deduction 9, the concern is with a time interval spent within the alternative. These distinctions can be kept clear in

[5] Strictly speaking, Deductions 6*a*, 6*b*, 15 hold only for situations in which there is empirical support for the stated relationship between brain or receptor injury and discriminability.

[6] See footnote 2.

the formulation of a new deduction in terms of the interval between choices.

Deduction 9*a*. In the simple two-alternative situation, with the time interval between choices constant, the greater the proportion of time between choices spent in the presence of the chosen alternative, the greater the amount of spontaneous alternation on the succeeding trial.

Deductions 7 and 9 can now be combined in the following.

Deduction 10. In the simple two-alternative situation, if *S* is delayed between choices and this delay occurs within the alternative last chosen, *S* will show more spontaneous alternation than under a condition of no delay. If, however, *S* is delayed between choices and this delay occurs outside the alternative last chosen, *S* will show less spontaneous alternation than under a condition of no delay.

Verified; crucial.

The author (8) found that when rats were detained in the end box of the alternative chosen in the first run in a T maze, 96 per cent of the group showed spontaneous alternation on the following trial. When they were kept in other parts of the maze at the end of their first run, between 36 and 68 per cent of the group showed spontaneous alternation. When they were not detained, spontaneous alternation varied between 72 and 88 per cent.

Deduction 11. In the simple two-alternative situation, if *S* is given forced runs to one alternative, the greater the number of such forced trials, the greater the amount of spontaneous alternation on a subsequent free-choice trial.

Verified; not crucial.

This procedure is basically the same as that involved in Deduction 9. In Deduction 11 the animal builds up stimulus satiation during repeated visits to one alternative; in Deduction 9 it does so during a period of detention within the alternative. Zeaman and House, deriving their hypothesis from the reactive-inhibition postulate, verified Deduction 11.

The amount of spontaneous alternation obtained by each of the two procedures, by forced trials and by detention at the alternative, may now be compared to yield a clearly critical deduction.

Deduction 12. In the simple two-alternative situation, detention of *S* for a given period of time at the alternative last chosen will yield larger amounts of spontaneous alternation than will a series of forced trials to one alternative over the same period of time.

Not tested; crucial.

In the forced-trial procedure, as compared with the detention procedure, *S*, being taken from end box to starting box, spends a smaller proportion of the time within the alternative itself. Thus *S* has less time to develop stimulus satiation.

Variability (*Preference for Variable Situations*)

An important variation of the two-alternative situation is one in which the cues of one alternative are varied while the cues of the other are kept constant. This type of situation has been investigated in two studies of "variability." Krechevsky (13, 14) used a two-alternative maze, both arms of which contained a number of turns. One arm could be varied in pattern of turns from trial to trial. Both variable and constant arms were constructed so that there was an equal number of right and left turns within each arm.

Three different maze arrangements were used: a long varying versus a short constant path (13); a short varying versus a long constant path (14); a long constant versus a short constant path (14). In each of these Krechevsky ran two groups of rats, one group normal and the other brain-operated.

In such a situation, stimulus satiation to the varying alternative will be reduced by its division among the different forms of the varying alternative. Essentially, a situation exists in which the constant alternative A is opposed by a family of alternatives B_1, B_2, B_3, . . . B_n. Alternative A bears the weight of all of the stimulus satiation built up during an exposure to it; the family of alternatives B divides stimulus satiation among its variations.

Corollary III. If an alternative is varied from trial to trial, then less stimulus satiation will be built up to it than if no such variation occurs.

Deduction 13. In the simple two-alternative situation, if one of the alternatives is varied from trial to trial and the other is kept constant, then *S* will prefer the varied alternative.

Verified; crucial.

This is verified by Krechevsky (13, 14). Examination of the results for his normal animals shows that the predicted preference is strong enough to cause *S*s to prefer the varied path even when it is the longer path.[7]

It follows that the division of stimulus satiation among the variations of the changing alternative depends on the degree of discriminability of the variations. As the variations become less and less discriminable, the condition of the constant alternative, i.e., of no change, is approached.

Corollary IV. If an alternative is varied from trial to trial, the more discriminable the variations the less the amount of stimulus satiation that will be present for the alternative.

[7] Krechevsky's findings cannot be explained in terms of reactive inhibition or any other theory in terms of the response since, as pointed out earlier, the mazes were designed so that each arm had the same number of right and left turns within it. The *S*s would have, following each trial, an equal amount of reactive inhibition to right and left turning.

There were indicated earlier two ways to reduce discriminability: reduction of the number of distinguishing cues of the perceived objects and impairment of the sensory or nervous systems of the *S*s. Consideration of these methods leads to the following deductions.

Deduction 14. In the simple two-alternative situation, the preference for a varying alternative (as opposed to a constant alternative) will be greater, the greater the difference between variations.

Not tested; crucial.

Deduction 15. In the simple two-alternative situation, brain-injured *S*s will show less preference for a varying alternative (as opposed to a constant alternative) than will normal *S*s.

Verified; crucial.

In Krechevsky's studies (13, 14) the normal animals preferred the varying alternative more strongly than did the operated animals.

Finally, it can be deduced that the greater the number of variations that share the stimulus satiation, the less the amount present to each.

Deduction 16. In the simple two-alternative situation, the greater the number of variations, the greater the preference for the varying alternative.

Not tested; crucial.

DEDUCTIONS FOR THE MULTIPLE-ALTERNATIVE SITUATION

We may now generalize the deductions to apply to situations with more than two alternatives. For example, let us apply the postulate to the behavior of *S* in the three-alternative situation. The *S*, after choosing one alternative, will build up stimulus satiation to it. Upon returning to the choice point for a second trial, *S* will probably choose a new alternative since responsiveness to the first choice has been reduced. If returned again to the

choice point, *S* will probably choose the "unsatiated" third alternative.

Deduction 1′. In a simple multiple-alternative situation, *S* will choose many alternatives in a series of consecutive trials.

Definition 3. Many alternatives mean more alternatives than could reasonably be expected on the basis of chance.

Deduction 1′ is verified by the findings of Wingfield and Dennis (22) on the number of different paths chosen by rats running four trials on a four-alternative maze.

Deduction 1 appears as simply a special case of Deduction 1′. In fact, a set of more general deductions can now be obtained in most cases merely by changing "two-alternative" to "multiple-alternative" and "spontaneous alternation" to "tendency to choose many alternatives." Each of the 16 earlier deductions becomes a special case of one of the new set of 16 deductions for multiple-alternative situations.

Consideration of situations involving more than two alternatives suggests a new variable, the number of alternatives, which can form the basis of several new deductions. One such deduction, of particular interest, concerns the cumulation effect of Deductions 3 and 5 and its counterpart in multiple-alternative situations, i.e., the decrease in the tendency to choose many alternatives in an extended series of trials.

Deduction 17′. If there are two simple multiple-alternative situations, and one of them has a greater number of alternatives than the other, then the cumulation effect will be smaller and will develop more slowly for the situation with a greater number of alternatives.

Cumulation effect depends on the summation of amounts of stimulus satiation from successive visits to the same alternative. In the two-alternative situation, *S* visits alternative A and then B, then A again, then B again, rapidly building up near-maximum amounts of stimulus satiation to both alternatives. According to Deduction 1′, the *S*, in a three-alternative situation, visits A then B then C and then A again, B again, and C again. The time between successive visits to the same alternative is longer, thereby delaying the onset of maximum stimulus satiation for any single alternative. It will consequently take more time to reach the period when differences in the amounts of stimulus satiation present for the various alternatives approach zero.

DEDUCTIONS FOR "FREE" PROCEDURE

Deductions have been drawn thus far only for the procedure in which *S* makes a choice and then is returned by *E* to the starting point for the next trial. Another procedure, which we shall call the "free" procedure, is one that allows *S* to make as many choices as it will for a given period of time. The *S* makes a choice and then returns itself to the choice point for its next choice. When employed without rewards, this is the standard procedure in the study of "exploratory behavior." [8]

Exploratory behavior may now be derived from the stimulus satiation postulate. The deductions can be further generalized to cover this more complex procedure so that exploratory behavior, like spontaneous alternation, becomes a special case governed by the postulate. The word "choice" replaces the word "trial" in the deductions, since we are no longer concerned with a trial as defined earlier; and the procedure is no longer called "simple."

Most of the deductions have not been tested in the "free" procedure. Those

[8] The relationship between spontaneous alternation and exploratory behavior was pointed out by Tolman (20) who explained the former in terms of the latter. We, however, view both spontaneous alternation and exploratory behavior as manifestations of a single underlying factor.

for which there are relevant data have been verified. For example, there is the following deduction involving cumulation effect.

Deduction 3″. In the multiple-alternative situation, if an extended series of consecutive choices is given, as the number of such choices increases, the tendency to choose many alternatives decreases.

This has been verified by Dennis and Sollenberger (6) and later by Montgomery (17). Both studies found that rats enter fewer and fewer alleys in successive periods of free exploration.

Another deduction for which there are data is the following:

Deduction 17″. If there are two multiple-alternative situations and one of them has a greater number of alternatives than the other, then the cumulation effect will be smaller and develop more slowly for the situation with a greater number of alternatives.

This has also been verified by Dennis and Sollenberger (6) and Montgomery (17). They found that larger mazes elicited more activity and continued to elicit activity over a longer period of time than did small mazes. Larger mazes are considered here to contain a greater number of alternatives.

POSSIBLE APPLICATIONS OF THE THEORY TO HUMAN SUBJECTS

Although most of the work in the areas under consideration has been carried out with rats as *Ss*, there is some evidence obtained from human *Ss* that is congruent with the postulate.

The closest analogue to the two-alternative situation that elicits spontaneous alternation in rats is found in a study by Wingfield (21) who required college students to turn on one of a pair of lights. In a series of four trials the *Ss* not only spontaneously alternated the lights they chose to turn on (Deduction 1) but also showed more spontaneous alternation on the first pair than on the second pair of trials (Deduction 3) and more with lights of different hue than with lights of the same hue (Deduction 5).

Karsten's (12) work on the phenomenon which she calls "satiation" (Sättigung) is of particular interest. She had her *Ss* perform repeatedly such tasks as drawing lines or tapping.[9] This activity produces repeated experience of certain stimulus-objects, e.g., lines drawn, and can therefore be considered a case covered by Deduction 13″. This deduction can be considered to predict Karsten's finding of the *Ss*' resistance to continuing the task as well as their variations from the prescribed activity.[10] The phenomenon of cosatiation (transfer of satiation symptoms to new and similar tasks) could also be predicted from part iii of the postulate.

A further finding by Karsten suggests the relevance of the postulate to the problem of fatigue. Symptoms of satiation, including the *Ss*' inability to move their arms, disappeared when the grouping of the lines being drawn was changed even though the required muscle movements remained the same. This recalls two things: part iii of the postulate and the recent emphasis on factors other than simple muscle states in understanding the phenomenon of fatigue (1).

SUMMARY

Because the reactive-inhibition explanation of spontaneous alternation in rats seems inadequate, a new theory in terms of the effects of continued exposure to stimuli is proposed. This is

[9] Berlyne (3) handles these data in the same way in his expansion of Hull's theory to include perceptual "responses."

[10] The *Ss*' resistance also expressed itself in the form of inattentiveness and "fatigue." All of these symptoms could be predicted on the basis of Corollary I.

presented in the form of a postulate with deductions drawn for various aspects of the simple two-alternative situation and available evidence for their validity presented. Following this, the theory is generalized to cover situations with any number of alternatives, and to situations that require other procedures, including one that elicits exploratory behavior. Finally the pertinence of the theory to some aspects of human behavior is indicated. The postulate of stimulus satiation yields novel predictions and serves to unify under a single explanatory construct such diverse phenomena as spontaneous alternation, exploratory behavior, Krechevsky's "variability," and Karsten's "satiation."

REFERENCES

1. BARTLEY, S. H. Fatigue and efficiency. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: Van Nostrand, 1951. Pp. 318–348.
2. BERLYNE, D. E. Novelty and curiosity as determinants of exploratory behavior. *Brit. J. Psychol.*, 1950, **41**, 68–80.
3. BERLYNE, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
4. DENNIS, W. A comparison of the rat's first and second exploration of a maze unit. *Amer. J. Psychol.*, 1935, **47**, 488–490.
5. DENNIS, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, **28**, 305–312.
6. DENNIS, W., & SOLLENBERGER, R. J. Negative adaptation in the maze exploration of albino rats. *J. comp. Psychol.*, 1934, **18**, 197–206.
7. GLANZER, M. Stimulus satiation as an explanation of spontaneous alternation in rats. Unpublished Ph.D. dissertation, Univer. of Michigan, 1952.
8. GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.
9. HEATHERS, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10**, 359–380.
10. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
11. JACKSON, M. M. Reactive tendencies of the white rat in running and jumping situations. *J. comp. Psychol.*, 1941, **31**, 255–262.
12. KARSTEN, A. Untersuchungen zur Handlungs- und Affektpsychologie: V. Psychische "Sattigung." *Psychol. Forsch.*, 1928, **10**, 142–154.
13. KRECHEVSKY, I. Brain mechanisms and variability: II. Variability where no learning is involved. *J. comp. Psychol.*, 1937, **23**, 139–163.
14. KRECHEVSKY, I. Brain mechanisms and variability: III. Limitations of the effect of cortical injury upon variability. *J. comp. Psychol.*, 1937, **23**, 351–364.
15. LASHLEY, K. S. The mechanism of vision. II. The influence of cerebral lesions upon the threshold of discrimination for brightness in the rat. *J. genet. Psychol.*, 1930, **37**, 461–480.
16. MONTGOMERY, K. C. Spontaneous alternation as a function of time between trials and amount of work. *J. exp. Psychol.*, 1951, **42**, 82–93.
17. MONTGOMERY, K. C. Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.*, 1952, **45**, 50–57.
18. MONTGOMERY, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45**, 287–294.
19. SOLOMON, R. L. The role of effort in the production of several related behavior phenomena. Unpublished Ph.D. dissertation, Brown Univer., 1947.
20. TOLMAN, E. C. Purpose and cognition: The determiners of animal learning. *Psychol. Rev.*, 1925, **32**, 285–297.
21. WINGFIELD, R. C. Some factors influencing spontaneous alternation in human subjects. *J. comp. Psychol.*, 1943, **35**, 237–244.
22. WINGFIELD, R. C., & DENNIS, W. The dependency of the rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, **18**, 135–147.
23. ZEAMAN, D., & HOUSE, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41**, 177–186.

[MS. received July 2, 1952]

Psychological Review
Vol. 60, No. 4, 1953

THE BACKWARD CURVE: A METHOD FOR THE STUDY OF LEARNING

KEITH J. HAYES

Yerkes Laboratories of Primate Biology, Orange Park, Florida

Graphical representation of the temporal course of improvement in performance has long been a popular research tool in the field of learning theory. Such curves have often been presented as evidence for basic mechanisms presumed to be involved in learning. Potentially, this type of analysis is important: If it can be shown that the acquisition of simple associations is practically always a sigmoid (or linear, or discontinuous, etc.) function of training, then one theory may gain considerably more than another in presumptive validity.

Unfortunately, the type of curve commonly used for this purpose indicates the average performance of a group of *Ss* at successive stages of practice. These curves are quite irrelevant to basic problems of learning theory, since their forms are determined not only by the forms of their component individual curves, but also by the distributions of the individual curves. The influence of this second factor is illustrated in Fig. 1, where it is seen that sigmoid average curves tend to result from symmetrical distributions of various kinds of individual curves.[1] When the distribution of individual curves has a strong positive skew, the resulting average curve tends toward predominantly negative acceleration.

The only curves which bear directly on basic theoretical problems are those which represent individual instances of learning—single associations acquired by individual subjects. However, it is difficult to obtain individual curves for the simpler learning situations. Complex tasks such as motor skills, long mazes, or difficult syllable lists may give reliable individual curves, but they are not directly applicable to the more basic problems. Puzzle boxes, discriminations, and conditioned responses are learned so quickly that we can rarely get as many as ten reliable samples of performance during the course of training. Group curves are acceptable substitutes only when we can assume that they are representative of their component individual curves, and this assumption is seldom justified.

Failure of an average curve to represent individual learning is due to individual differences among the subjects. If the only differences are in speed of learning, the general form of the learning function can be approximated reasonably well by a Vincent curve. However, we can seldom justify the assumption that individual curves do not also differ in form. Even if homogeneity of form exists, the irregularity of individual curves makes it difficult to demonstrate. In some cases, form may be systematically related to learning speed, and Hilgard (2, p. 293) has suggested that separate Vincent curves should then be drawn for fast, slow, and perhaps intermediate learning—which is

[1] Normally distributed sigmoid individual curves will give a *similar* average curve; but Merrell (3) has shown (in the context of logistic growth curves) that the average curve need not have the *same* form as its components. The average curve may not be expressible as $y = K/1 + e^{a+bx}$, for instance, even though all of the individual curves are. Sidman (5) has made this same point with reference to the negatively accelerated function, $y = M - Me^{-bx}$.

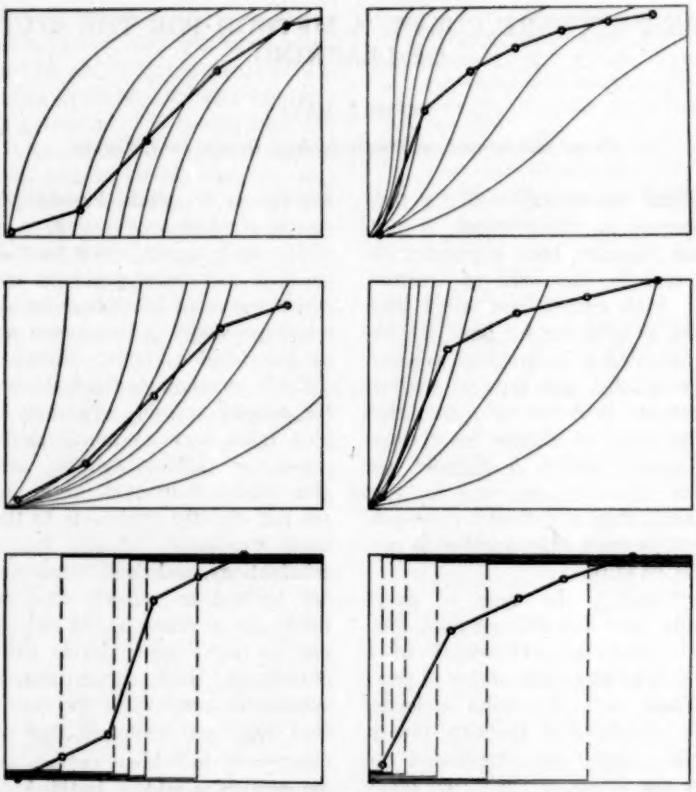

FIG. 1. The effect of distribution on the forms of average learning curves. Light lines show hypothetical individual curves; heavy lines show the resulting average curves. Broken vertical lines indicate discontinuity.

practical only if an adequate number of subjects is available in each of the necessary speed categories.

Where generally heterogeneous groups are involved, it is apparently impossible to plot satisfactory average curves of the complete course of learning. However, certain theoretical issues can be dealt with by examining smaller segments of the learning process, and a method is available for plotting average curves of such segments.[2] The general features of this method may be illustrated by analysis of some data made available by Mr. Robert Thompson.[3] Forty rats were trained in a brightness discrimination problem, to a criterion of ten consecutive correct responses. Motivation was escape from water, and correction was allowed.

The traditional average curve for this experiment is shown in Fig. 2, where each point represents 400 choices—ten trials for each of 40 rats.[4] The light

[2] A procedure very similar to the one described here has been used by Shuttleworth (4) in his study of the prepubertal spurt in human growth.

[3] Formerly with the Yerkes Laboratories, now at the University of Texas.

[4] Means are based on the assumption that no more errors would be made after achieve-

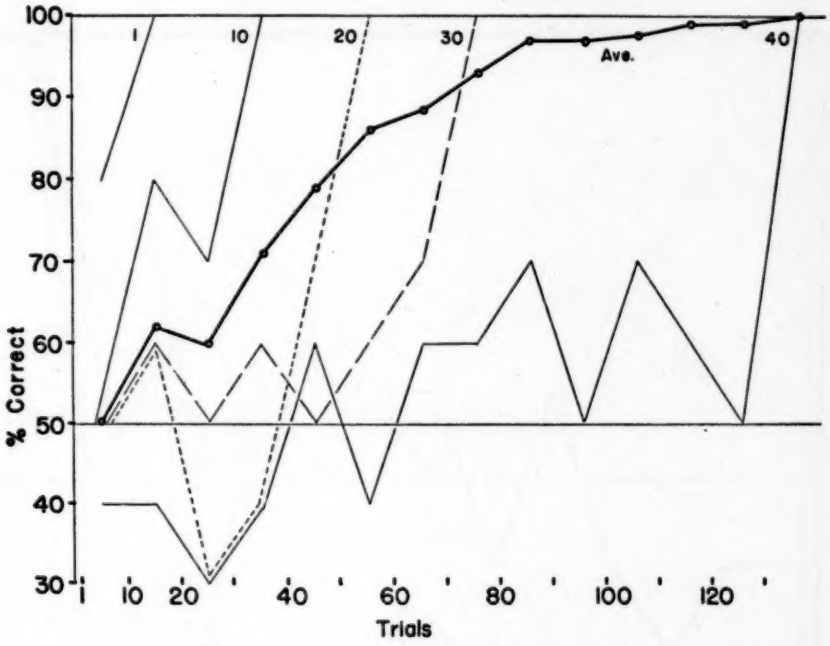

FIG. 2. Brightness discrimination learning in rats. The heavy line is the average learning curve ($N = 40$); light lines are individual curves for the 1st, 10th, 20th, 30th, and 40th rats to reach criterion.

lines in the same figure are individual learning curves for the 1st, 10th, 20th, 30th, and 40th animals to reach criterion. When due allowance has been made for the irregularity of the individual curves, the impression remains that the average curve has grossly distorted the individual data. In contrast to the gradual, negatively accelerated, group curve, the individual curves suggest a variable period of no progress, followed by uniformly sudden (or "insightful") learning.

The form of these individual curves raises the problem of how to plot an average curve which is capable of showing such an abrupt rise to mastery. The proposed solution is shown graphically in Fig. 3. Here the individual curves are displaced horizontally, so that their final points coincide. It is now possible to draw a single curve which does no serious injustice to any of the individuals. Such a curve will not tell us the average accuracy of performance on the x^{th} trial, but it will indicate the average accuracy x trials before achieving the criterion.

The actual construction of this curve involves plotting percentages for successive trials, not working forward from Trial 1, but rather working backward from the criterion. The heavy line in Fig. 4 is such a *backward learning curve*, each point representing the percentage correct on a single trial, for 40 rats.

Backward curves, like Vincent curves, should not be drawn through points based on the criterial trials or the im-

ment of criterion. The "true" curve would be slightly lower, except near the origin.

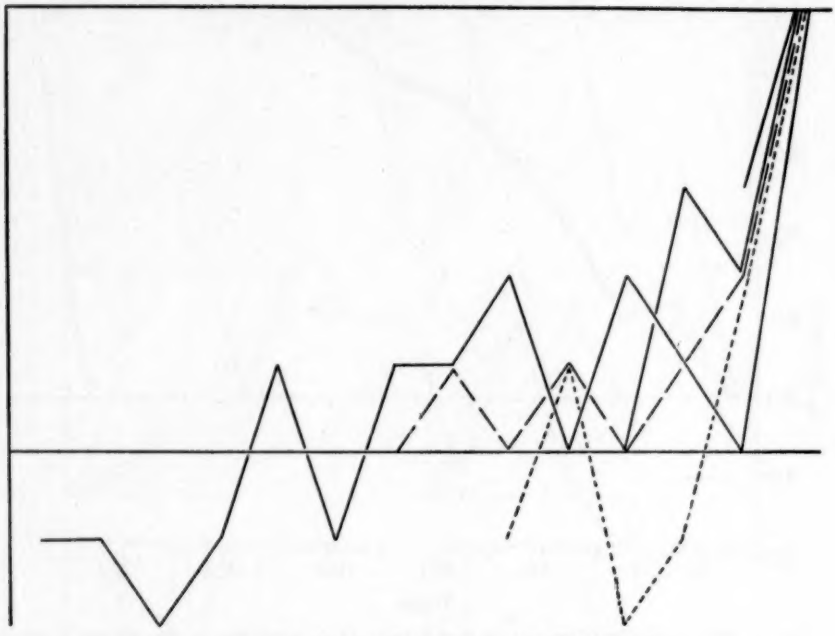

FIG. 3. The individual curves of Fig. 2, shifted horizontally

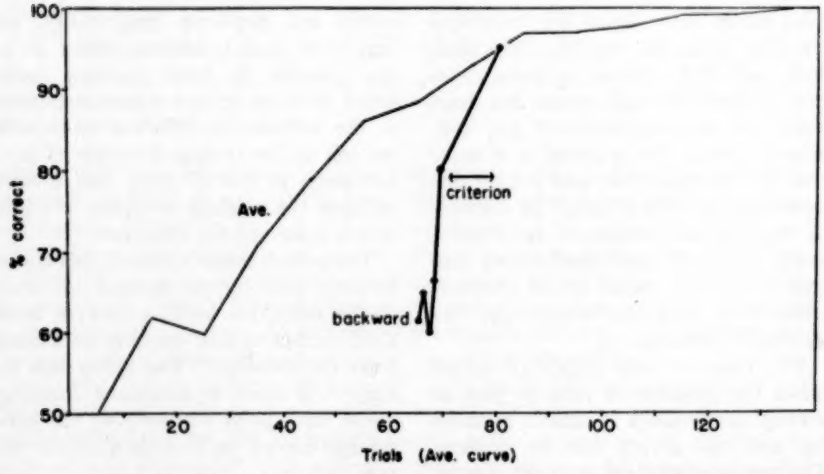

FIG. 4. Comparison of the traditional average curve with the backward curve. The horizontal position of the backward curve was arbitrarily chosen to make the two curves coincide at the 95% level.

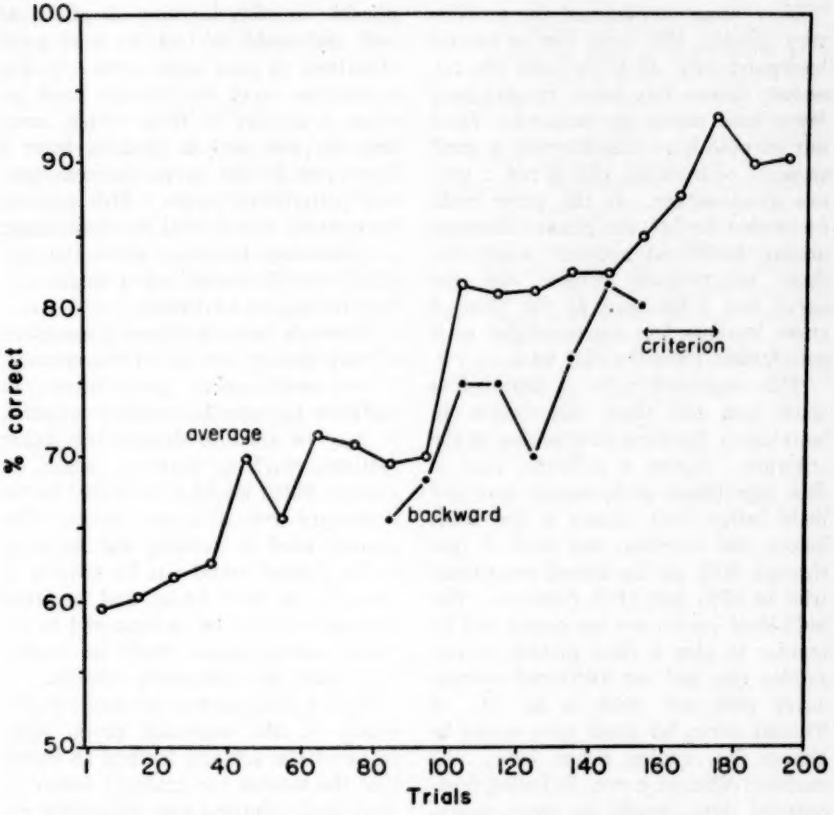

FIG. 5. Average and backward curves for 23 cases of visual discrimination learning by monkeys. Points are based on successive groups of 10 trials. From 90 to 392 trials were required to reach a criterion of 20 successive correct choices. (Data supplied by K. L. Chow.)

mediately preceding trial (2, p. 286).[5] Because of this restriction, it is desirable to have one or more postcriterial trials, to represent the final stage of learning. A postcriterial point was provided in this case by reanalyzing the data in terms of a criterion of nine successive correct trials, instead of the ten originally used in training. (If the rats had been trained with this analysis in mind, they would have been run for several trials beyond criterion, so that postcriterial slope, as well as level, could have been determined.)

[5] Attainment of criterion is due primarily to the subject's increased ability, but it is also due in part to chance fluctuation—which the criterion catches on the upswing. If the criterion is not stringent enough, chance may contribute more than ability. Grant (1) has prepared tables from which appropriate criteria may be selected for discrimination training.

If criterial trials are admitted in a backward curve, "insightful learning" can be demonstrated in a group of tossed coins. This may take time if a stringent criterion is used, but most coins will eventually choose heads ten times in a row if given enough training. Coins differ from other species in showing no retention of the habit on postcriterial trials.

Since some rats learned the problem very quickly, the curve can be carried backward only six trials from the criterion; before this point, progressively fewer rats would be included. Since our purpose is to examine only a small segment of learning, this is not a serious disadvantage. If the curve could be carried further, the phase differences among individual subjects would become progressively larger, and the curve less valid—just as the forward curve becomes less representative as it gets farther from the first trial.

This backward curve is intended to show just one thing—the course of learning in the immediate region of the criterion. Figure 4 indicates that in this experiment performance averaged little better than chance a few trials before the criterion, and that it rose through 80% on the second precriterial trial to 95% just after criterion. The individual curves are too coarse and irregular to give a clear picture of this sudden rise, and the traditional average curve does not show it at all. A Vincent curve for these data would be almost flat, ending below 70%. (A modified Vincent curve, including postcriterial data, would be more appropriate.)

Group data are commonly so heterogeneous that the customary type of average curve is bound to give an impression of gradual learning, regardless of the actual facts. The question may arise as to whether the backward curve is foredoomed to give an impression of sudden learning. The answer is illustrated in Fig. 5, which shows forward and backward curves of equal slope, both based on the same data. The backward curve does demonstrate gradual learning, when it occurs.

Discrimination learning has been used in these examples only because discrimination data happened to be available. The general method is applicable to other learning situations as well and could, in fact, be used more effectively in some other cases. In discrimination work the criterion must include a number of trials which, since they are not used in plotting, leave a large gap in the curve, between pre- and postcriterial points. This gap can be reduced to one trial in conditioning or puzzle-box learning, where the amplitude or time score for a single trial may be used as a criterion.

Although examples have been given of only one application of this method, it has more general possibilities. In addition to detecting sudden solution, it may be used to demonstrate other features (such as plateaus, peaks, or spurts) which would be obscured in the customary over-all group curves. The criteria used in defining the segments to be plotted would not be criteria of mastery, in such cases, and practical difficulties might be encountered in devising criteria which would be reasonably short, yet adequately reliable.

Special analysis was not entirely necessary in the examples given, since many of the subjects learned so slowly that the sudden (or gradual) nature of their final solutions was reasonably apparent in the individual curves. The advantage of the method would be greater in cases of faster learning, where usable individual curves cannot be obtained. It has some utility even with slow learning, however, since large numbers of individual curves are cumbersome to deal with, and impractical to publish.

Speculation as to why the rats of Fig. 4 learned so much more suddenly than the monkeys of Fig. 5 would be hazardous, since the two experiments differed in many ways. It may eventually be profitable to look for the cause of such differences, but only when more cases of sudden learning are available. It is hoped that the procedure sug-

gested here may make some such cases manifest, which might otherwise remain latent in the form of inadequately analyzed data.

SUMMARY

A method is proposed for plotting small segments of group learning curves in such a way that details are shown which would be obscured in the customary types of average curves.

The method is illustrated by search for sudden solution in two batches of discrimination data. Sudden learning was found in one case, and not in the other.

Suggestions are offered for other applications of the method.

REFERENCES

1. GRANT, D. A. Additional tables of the probability of "runs" of correct responses in learning and problem solving. *Psychol. Bull.*, 1947, **44**, 276–279.
2. HILGARD, E. R. A summary and evaluation of alternative procedures for the construction of Vincent curves. *Psychol. Bull.*, 1938, **35**, 282–297.
3. MERRELL, MARGARET. The relationship of individual growth to average growth. *Human Biol.*, 1931, **3**, 37–70.
4. SHUTTLEWORTH, F. K. The physical and mental growth of girls and boys age six to nineteen in relation to age at maximum growth. *Monogr. Soc. Res. Child Develpm.*, 1939, **4**, No. 3.
5. SIDMAN, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, **49**, 263–269.

[MS. received July 21, 1952]

Psychological Review
Vol. 60, No. 4, 1953

A THEORY OF STIMULUS VARIABILITY IN LEARNING[1]

W. K. ESTES AND C. J. BURKE

Indiana University

There are a number of aspects of the stimulating situation in learning experiments that are recognized as important by theorists of otherwise diverse viewpoints but which require explicit representation in a formal model for effective utilization. One may find, for example, in the writings of Skinner, Hull, and Guthrie clear recognition of the statistical character of the stimulus concept. All conceive a stimulating situation as made up of many components which vary more or less independently. From this locus of agreement, strategies diverge. Skinner (17) incorporates the notion of variability into his stimulus-class concept, but makes little use of it in treating data. Hull states the concept of multiple components explicitly (13) but proceeds to write postulates concerning the conditions of learning in terms of single components, leaving a gap between the formal theory and experimentally defined variables. Guthrie (11) gives verbal interpretations of various phenomena, e.g., effects of repetition, in terms of stimulus variability; these interpretations generally appear plausible but they have not gained wide acceptance among investigators of learning, possibly because Guthrie's assumptions have not been formalized in a way that would make them easily used by others. Statistical theories of learning differ from Hull in making stimulus variability a central concept to be used for explanatory purposes rather than treating it as a source of error, and they go beyond Skinner and Guthrie in attempting to construct a formalism that will permit unambiguous statements of assumptions about stimulus variables and rigorous derivation of the consequences of these assumptions.

It has been shown in a previous paper (7) that several quantitative aspects of learning, for example the exponential curve of habit growth regularly obtained in certain conditioning experiments, follow as consequences of statistical assumptions and need not be accounted for by independent postulates. All of the derivations were carried out, however, under the simplifying assumption that all components of a stimulating situation are equally likely to occur on any trial. By removing that restriction, we are now in a position to generalize and extend the theory in several respects. It will be possible to show that regardless of whether assumptions as to the necessary conditions for learning are drawn from contiguity theories or from reinforcement theories, certain characteristics of the learning process are invariant with respect to stimulus properties while other characteristics depend in specific ways upon the nature of the stimulating situation.

The Generalized Set Model: Assumptions and Notation

The exposure of an organism to a stimulating situation determines a set

[1] This paper is based upon a paper reported by the writers at the Boston meetings of the Institute of Mathematical Statistics in December 1951. The writers' thinking along these and related lines has been stimulated and their research has been facilitated by participation in an interuniversity seminar in mathematical models for behavior theory which met at Tufts College during the summer of 1951 and was sponsored by SSRC.

of events referred to collectively as stimulation. These events constitute the data of the various special disciplines concerned with vision, audition, etc. We wish to formulate our model of the stimulus situation so that information from these special disciplines can be fed into the theory, although utilization of that information will depend upon the demands of learning experiments.

For the present we shall make only the following very general assumptions about the stimulating situation: (*a*) The effect of a stimulus situation upon an organism may be regarded as made up of many component events. (*b*) When a situation is repeated on a series of trials, any one of these component stimulus events may occur on some trials and fail to occur on others; as a first approximation, at least, the relative frequencies of the various stimulus events when the same situation (as defined experimentally) occurs on a series of trials, may be represented by independent probabilities. We formulate these assumptions conceptually as follows:

(*a*) With any given organism we associate a set S^* of N^* elements.[2] The N^* elements of S^* are to represent all of the stimulus events that can occur in that organism in any situation whatever with each of these possible events corresponding to an element of the set. (*b*) For any reproducible stimulating situation we assume a distribution of values of the parameter θ; we represent by θ_i the probability that the stimulus event corresponding to the i^{th} element of S^* occurs on any given trial.

[2] In the sequel, various sets will be designated by the letter S, accompanied by appropriate subscripts and superscripts. The letter N, with the same arrangement of subscripts and superscripts, always denotes the size of the set.

It is assumed that any change in the situation (and we shall attempt to deal only with controlled changes corresponding to manipulations of experimental variables) determines a new distribution of values of the θ_i. By repeating the "same" situation, we mean the same as described in physical terms, and we recognize that, strictly speaking, repetition of the same situation refers to an idealized state of affairs which can be approached by increasing experimental control but possibly never completely realized.

It is recognized that some sources of stimulation are internal to the organism. This means that in order to have a reproducible situation in a learning experiment it is necessary to control the maintenance schedule of the organism and also activities immediately preceding the trial. In the present paper we shall not use the term "trial" in a sufficiently extended sense to necessitate including in the θ distribution movement-produced-stimulation arising from the responses occurring on the trial.

We have noted that the behavior on a given trial is assumed to be a function of the stimulus elements which are sampled on that trial. If in a given situation certain elements of S^* have a probability $\theta = 0$ of being sampled, those elements have a negligible effect upon the behavior in that situation. For this reason we often represent a specific situation by means of a reduced set S. An element of S^* is in S if and only if it has a non-zero value of θ in the given situation. These sets are represented in Fig. 1. In this connection, we must note that a probability of zero for a given event does not mean that the event can never occur "accidentally"; this probability has the weaker meaning that the relative frequency of occurrence of the event is zero in the long run. For a

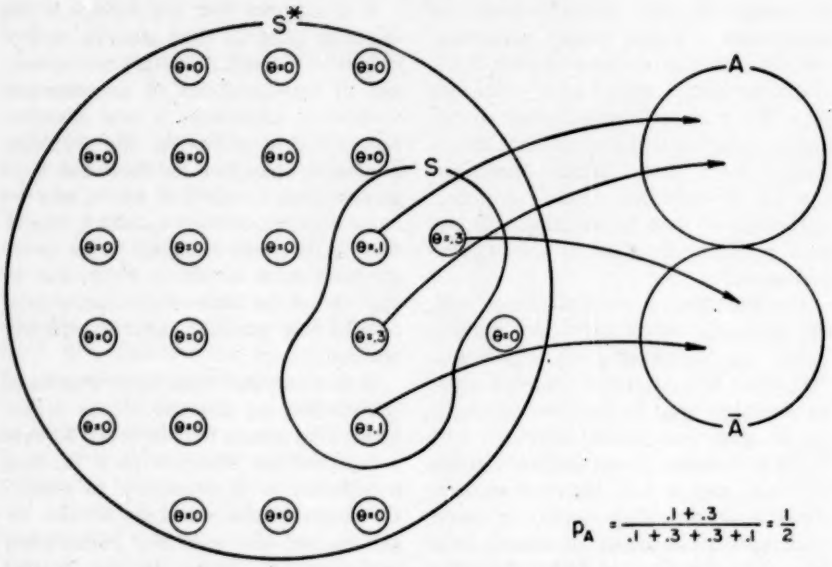

FIG. 1. A schematic representation of stimulus elements, the stimulus space S^*, the reduced set S containing elements with non-zero θ values for a given stimulating situation, and the response classes A and $\bar{A}$. The arrows joining elements of S to the response classes represent the partition of S into S_A and $S_{\bar{A}}$

more detailed explication of this point the reader is referred to Cramér (5).

It should be clearly understood that the probability, θ, that a given stimulus event occurs on a trial may depend upon many different environmental events. For example, a stimulus event associated with visual stimulation may depend for its probability upon several different light sources in the environment. Suppose that for a given stimulus element, the associated probability θ in a given situation depends only upon two separately manipulable components of the environment, a and b, and that the probabilities of the element's being drawn if only a or b alone were present are θ_a and θ_b, respectively. Then the probability attached to this element in the situation with both components present will be

$$\theta = \theta_a + \theta_b - \theta_a\theta_b.$$

THE RESPONSE MODEL

The response model formulated in a previous paper (7) will be used here without any important modification. We shall deal only with the simple case of two mutually exclusive and exhaustive response classes. The response class being recorded in a given situation will be designated A and the complementary class, $\bar{A}$. The dependent variable of the theory is the probability that the response occurring on a given trial is a member of class A. It is recognized that in a learning experiment the behaviors available to the organism may be classified in many different ways, depending upon the interests of the experimenter. The response class selected for investigation may be anything from the simplest reflex to a complex chain of behaviors involving many different groups of effectors. Adequate

treatment of all levels of response specification would require the formulation of a model for the response space and will not be attempted in the present paper. Preliminary investigation of this problem leads us to believe that when a response model is elaborated, the theory developed in this paper will be found to hold rigorously for the most elementary response components and to a first approximation for simple response classes that do not involve chaining. For experimental verification of the present theory we shall look to experiments involving response classes no more complex than flexing a limb, depressing a bar, or moving a key.

CONDITIONAL RELATIONS AND RESPONSE PROBABILITY

We assume that the behavior of an organism on any trial is a function, not of the entire population of possible stimulus events, but only of those stimulus events which occur on that trial; further, when learning takes place, it involves a change in the dependency of the response upon the stimulus events which have occurred on the given trial.

Conditional relations, or for brevity, connections, between response classes and stimulus elements are defined as in other papers on statistical learning theory (3, 7). The response classes A and $\bar{A}$ define a partition of S^* into two subsets $S_A{}^*$ and $S_{\bar{A}}{}^*$. Elements in $S_A{}^*$ are said to be "connected to" or "conditioned to" response A; those in $S_{\bar{A}}{}^*$ to response $\bar{A}$. The concept of a partition implies specifically that every element of S^* must be connected either to A or to $\bar{A}$ but that no element may be connected to both simultaneously.[3] Various features of the model are illustrated in Fig. 1.

For each element in S^* we define a quantity $F_i(n)$ representing the probability that the element in question is connected to response A, i.e., is in the subset $S_A{}^*$, at the end of trial n. The mean value of $F_i(n)$ over S^* is, then, simply the expected proportion of elements connected to A, and if all of the θ_i were equal, it would be natural to define this proportion as the probability, $p(n)$, that response A occurs on trial $n+1$. In the general case, however, not all of the θ_i are equal and the contribution of each element should be weighted by its probability of occurrence, giving

$$(1)\quad p(n) = \frac{\sum_i \theta_i F_i(n)}{\sum_i \theta_i} = \frac{1}{N^*\bar{\theta}} \sum_i \theta_i F_i(n).$$

It will be seen that in the equal θ case, expression (1) reduces to

$$(2)\quad p(n) = \frac{\theta}{N^*\theta} \sum_i F_i(n) = E(F_i(n))$$

which, except for changes in notation, is the definition used in previous papers (6, 7).

The quantity p is, then, another of the principal constructs of the theory. It is referred to as a probability, firstly because we assume explicitly that quantities p are to be manipulated mathematically in accordance with the axioms of probability theory, and secondly because in some situations p can be given a frequency interpretation. In any situation where a sequence of responses can be obtained under conditions of negligible learning and independent trials (as at the asymptote of a simple learning experiment carried out with discrete, well-spaced trials) the numerical value of p is taken as the average relative frequency of response A. For all situations the construct p is assumed to

[3] The argument of this section could as well be given in terms of the set S as of S^*, defining S_A and $S_{\bar{A}}$ as the partition of S imposed by the response classes A and $\bar{A}$.

correspond to a parameter of the behavior system, and we do not cease to speak of this as a probability in the case of a situation where it cannot be evaluated as a relative frequency. It has been shown in a previous paper (7) that p can be related in a simple manner to rate or latency of responding in many situations; thus in all applications of the theory, p is evaluated in accordance with the rules prescribed by the theory, either from frequency data or from other appropriate data, and once evaluated is treated for all mathematical purposes as a probability.

REPRESENTATION OF LEARNING PROCESSES

In order to account for the gradual course of learning in most situations, a number of the earlier quantitative theories, e.g., those of Hull (13), Gulliksen and Wolfle (10), Thurstone (18) have assumed that individual connections are formed gradually over a series of learning trials. Once we adopt a statistical view of the stimulating situation, however, it can be shown rigorously that not only the gradual course of learning but the form of the typical learning curve can be accounted for in terms of probability considerations even if we assume that connections are formed on an all-or-none basis. This being the case, there seems to be no evidence whatsoever that would require a postulate of gradual formation of individual connections. Psychologically an all-or-none assumption has the advantage of enabling us to account readily for the fact that learning is sudden in some situations and gradual in others; mathematically, it has the advantage of great simplicity. For these reasons, recent statistical theories of learning have adopted some form of the all-or-none assumption (3, 7, 15).

Under an all-or-none theory, we must specify the probabilities that any stimulus element that is sampled on a given trial will become connected to A or to $\bar{A}$. For convenience in exposition, we shall limit ourselves in this paper to the simplest special case, i.e., a homogeneous series of discrete trials with probability equal to one that all elements occurring on a trial become connected to response A.

We begin by asking what can be said about the course of learning during a sequence of trials regardless of the distribution of stimulus events. It will be shown that our general assumptions define a family of mathematical operators describing learning during any prescribed sequence of trials, the member of the family applicable in a given situation depending upon the θ distribution. We shall first inquire into the characteristics common to all members of a family, and then into the conditions under which the operators can be approximated adequately by the relatively simple functions that have been found convenient for representing learning data in previous work.

Let us consider the course of learning during a sequence of trials in the simplified situation. Each trial in the series is to begin with the presentation of a certain stimulus complex. This situation defines a distribution of θ over S^* so that each element in S^* has some probability, θ_i, of occurring on any trial, and we represent by S the subset of elements with non-zero θ values; any element that occurs on a trial becomes connected to A (or remains connected to A if it has been drawn on a previous trial). For concreteness the reader might think of a simple conditioning experiment with the CS preceding the US by an optimal interval, and with conditions arranged so that the UR is evoked on each trial and decremental factors are negligible; the situation represented by S is that obtaining from the

onset of the CS to the onset of the US, and the response probability p will refer to the probability of A in this situation. The number of elements in S will be designated by N. For simplicity we shall suppose in the following derivations that none of the elements in S are connected to A at the beginning of the experiment. This means that the learning curves obtained all begin with N_A and p equal to zero. No loss of generality is involved in this simplification; our results may easily be extended to the case of any arbitrary initial condition.

The i^{th} element in S will still remain in $S_{\bar{A}}$ after the n^{th} trial if and only if it is not sampled on any of the first n trials; the likelihood that this occurs is $(1-\theta_i)^n$. Hence, if $F_i(n)$ represents the expected probability that this element is connected to A after the n^{th} trial, we obtain:

$$(3) \qquad F_i(n) = 1 - (1-\theta_i)^n.$$

The expected number of elements in S connected to A after the n^{th} trial, $E[N_A(n)]$, will be the sum of these expected contributions from individual elements:

$$(4) \quad E[N_A(n)] = \sum_i F_i(n) = \sum_i [1-(1-\theta_i)^n] = N - \sum_i (1-\theta_i)^n.$$

We are now in a position to express p, the probability of response A, as a function of the number of trials in this situation. By substituting for the term $F_i(n)$ of equation (1) its equivalent from equation (3), we obtain the relation

$$(5) \quad p(n) = \frac{1}{N\bar{\theta}} \sum_i \theta_i [1-(1-\theta_i)^n] = 1 - \frac{1}{N\bar{\theta}} \sum_i \theta_i (1-\theta_i)^n.$$

Equation (5) defines a family of learning curves, one for each possible θ distribution, and it has a number of simple properties that are independent of the distribution of the θ_i. It can easily be verified by substitution that there is a fixed point at $p = 1$, and this will be the asymptote approached by the curve of $p(n)$ vs. n as n increases over all bounds. Members of the family will be monotonically increasing, negatively accelerated curves, approaching a simple negative growth function as the θ_i tend toward equality. If all of the θ_i are equal to $\bar{\theta}$, equation (5) reduces to

$$(6) \qquad p(n) = 1 - (1-\bar{\theta})^n$$

which, except for a change in notation, is the same function derived previously (7) for the equal θ case[4] and corresponds to the linear operator used by Bush and Mosteller (2) for situations where no decremental factor is involved. In mathematical form, equation (6) is the same as Hull's well-known expression for growth of habit strength, but the function does not have the same relation to observed probability of responding in Hull's theory as in the present formulation.

Except where the distribution function of the θ_i either is known, or can be assumed on theoretical grounds to be approximated by some simple expression, equation (5) will not be convenient to work with. In practice we are apt to assume equal θ_i and utilize equation (6) to describe experimental data. The nature of the error of approximation involved in doing this can be stated generally. Immediately after the first trial, the curve for the general case must lie above the curve for the

[4] This is essentially the same function developed for the equal θ case in a previous paper (7); the terms $\bar{\theta}$ and n of equation (6) correspond to the terms $q = s/S$, and T of that paper.

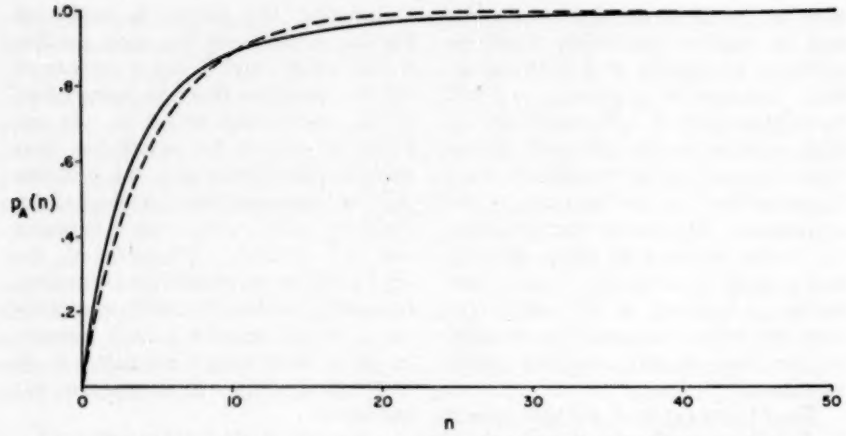

FIG. 2. Response probability, in S, as a function of number of trials for the numerical examples presented in the text. The solid curve is the exact solution for a population of elements, half of which have $\theta = 0.1$ and half $\theta = 0.3$. The dashed curve describes the equal θ approximation with $\bar{\theta} = 0.2$. Initially no elements of S are conditioned to A.

equal θ case; the difference between the two curves increases for a few trials, then decreases until they cross (in constructing hypothetical θ distributions of diverse forms we have usually found this crossing in the neighborhood of the fourth to eighth trial); after crossing, the curves diverge to a smaller extent than before, then come together as both go to the same asymptote at $p = 1$. It can be proved that the curves for the general and special case cross exactly once as n goes from one to infinity. We cannot make any general statement about the maximum error involved in approximating expression (5) with expression (6), but after studying a number of special cases, we are inclined to believe that the error introduced by the approximation will be too small to be readily detectable experimentally for most simple learning situations that do not involve compounding of stimuli.

The development of equations (5) and (6) has necessarily been given in rather general terms, and it may be helpful to illustrate some of the considerations involved by means of a simple numerical example. Imagine that we are dealing with some particular conditioning experiment in which the CS can be represented by a set S, composed of two subsets of stimulus elements, S_1 and S_2, of the sizes $N_1 = N_2 = N/2$, where N is the number of elements in S. Assume that for all elements in S_1 the probability of being drawn on any trial is $\theta_1 = 0.3$ and for those in S_2, $\theta_2 = 0.1$. Now we wish to compute the predicted learning curve during a series of trials on which A responses are reinforced, assuming that we begin with all elements connected to A. Equation (5) becomes

$$p(n) = 1 - \frac{1}{0.2N}[N_1(0.3) \times (1-0.3)^n + N_2(0.1)(1-0.1)^n]$$
$$= 1 - \frac{1}{0.4}[0.3(0.7)^n + 0.1(0.9)^n].$$

Plotting numerical values computed from this equation, we obtain the solid curve given in Fig. 2.

Now let us approach the same prob-

lem, but supposing this time that we know nothing about the different θ values in the subsets S_1 and S_2 and are given only that $\bar{\theta} = 0.2$. We now obtain predicted learning curves under the equal θ approximation. Equation (6) becomes:

$$p(n) = 1 - (1 - 0.2)^n$$

and numerical values computed from this yield the dashed curve of Fig. 2.

Inspection of Fig. 2 shows that the exact treatment leads to higher values of $p(n)$ on the early trials but to lower values on the later trials, the difference becoming negligible for large n. The reason, in brief, for the steeper curvature of the exact curve is that elements with high θ values are likely to be drawn, and therefore conditioned to A_1 earlier in the learning process than elements with low θ values, and then because they will tend to recur frequently in successive samples, to lead to relatively high values of p. During the late stages of learning, elements with low θ values that have not been drawn on the early trials will contribute more unconnected elements per trial than would be appearing at the same stage with an equal θ distribution and will depress the value of p below the curve for the equal θ approximation.

It should be emphasized that the generality of the present approach to learning theory lies in the concepts introduced and the methods developed for operating with them, not in the particular equations derived. Equation (5), for example, can be expected to apply only to an extremely narrow class of learning experiments. On the other hand, the methods utilized in deriving equation (5) are applicable to a wide variety of situations. For the interest of the experimentally oriented reader, we will indicate briefly a few of the most obvious extensions of the theory developed above, limiting ourselves to the equal θ case.

As written, equation (6) represents the predicted course of conditioning for a single organism with an initial response probability of zero. We can allow for the possibility that an experiment may begin at some value of $p(0)$ other than zero by rewriting (6) in the more general form

$$(7) \quad p(n) = 1 - [1 - p(0)](1 - \bar{\theta})^n$$

which has the same form as (6) except for the initial value.

If we wish to consider the mean course of conditioning in a group of m organisms with like values of $\bar{\theta}$ but varying initial response probabilities, we need simply sum equation (7) over the group and divide by m, obtaining

$$(8) \quad \bar{p}(n) = \frac{1}{m}\sum p(n) = 1 - [1 - \bar{p}(0)](1 - \bar{\theta})^n.$$

The standard deviation of $p(n)$ under these circumstances is simply

$$(9) \quad \sigma_p(n) = \sqrt{\frac{1}{m}\sum p^2(n) - \bar{p}^2(n)} = (1 - \bar{\theta})^n \sigma_p(0)$$

where $\sigma_p(0)$ is the dispersion of the initial p values for the group. Variability around the mean learning curve decreases to zero in a simple manner as learning progresses.

The treatment of counter-conditioning, i.e., extinguishing one response by giving uniform reinforcement to a competing response, follows automatically from our account of the acquisition process. Returning to equation (6) and recalling that the probabilities of A_1 and $\bar{A}_1$ must always sum to unity, we note that while response A_1 undergoes conditioning in accordance with (6), response $\bar{A}_1$ must undergo extinction in

accordance with the function

$$p_{\bar{A}}(n) = 1 - p_A(n) = (1 - \bar{\theta})^n.$$

If, then, we begin with any arbitrary $p(0)$ and arrange conditions so that $\bar{A}$ is evoked and conditioned to all elements drawn on each trial, the extinction of response A will be given by the simple decay function

$$p(n) = p(0)(1 - \bar{\theta})^n. \tag{10}$$

Again the mean and standard deviation of $p(n)$ can easily be computed for a group of organisms with like values of $\bar{\theta}$ but varying values of $p(0)$:

$$\bar{p}(n) = \bar{p}(0)(1 - \bar{\theta})^n \tag{11}$$

$$\sigma_p(n) = (1 - \bar{\theta})^n \sigma_p(0). \tag{12}$$

As in the case of acquisition, variability around the mean curve decreases to zero in a simple manner over a series of trials.

Since variability due to variation in $p(0)$ is reduced during both conditioning and counter-conditioning, it will be seen that in general we should expect less variability around a curve of relearning than around a curve of original learning for a given group of subjects.

Application of the Statistical Model to Learning Experiments

Since our concern in this paper has been with the development of a stimulus model of considerable generality, it has been necessary in the interests of clear exposition to omit reference to most of the empirical material upon which our theoretical assumptions are based. The evaluation of the model must rest upon detailed interpretation of specific experimental situations. It is clear, however, that the statistical model developed here cannot be tested in isolation; only when it is taken together with assumptions as to how learning occurs and with rules of correspondence between terms of the theory and experimental variables, will experimental evaluation be possible. Limitations of space preclude a detailed theoretical analysis of individual learning situations in this paper. In order to indicate how the model will be utilized and to suggest some of its explanatory potentialities we shall conclude with a few general remarks concerning the interpretation of learning phenomena within the theoretical framework we have developed.

Application of the model to any one isolated experiment will always involve an element of circularity, for information about a given θ distribution must be obtained from behavioral data. This circularity disappears as soon as data are available from a number of related experiments. The utility of the concept is expected to lie in the possibility of predicting a variety of facts once the parameters of the θ distribution have been evaluated for a situation. The methodology involved has been illustrated on a small scale by an experiment (6) in which the mean θ value for an operant conditioning situation was estimated from the acquisition curve of a bar-pressing habit and then utilized in predicting the course of acquisition of a second bar-pressing habit by the same animals under slightly modified conditions.

When the statistical model is taken together with an assumption of association by contiguity, we have the essentials of a theory of simple learning. The learning functions (5), (6), and (10) derived above should be expected to provide a description of the course of learning in certain elementary experiments in the areas of conditioning and verbal association. It must be emphasized, however, that these functions alone will not constitute an adequate theory of conditioning, for a number of relevant variables, especially those con-

trolling response decrement, have not been taken into account in our derivations. In conditioning experiments where decremental factors are minimized, there is considerable evidence (1, 4, 9, 14, 16) that the curve of conditioning has the principal properties of our equation (5) and can be well approximated by the equal θ case (7). The fact that functions derived from the model can be fitted to certain empirical curves is a desirable outcome, of course, but cannot be regarded as providing a very exacting test of the theory; probably any contemporary quantitative theory will manage to accomplish this much. On the other hand, the fact that the properties of our learning functions follow from the statistical nature of the stimulating situation is of some interest; in this respect the structure of the present theory is simpler than certain others, e.g., that of Hull (13), which require an independent postulate to account for the form of the conditioning curve.

It should also be noted that deviations from the exponential curve form may be as significant as instances of good fit. From the present model we must predict a specific kind of deviation when the stimulating situation contains elements of widely varying θ values. If, for example, curves of conditioning to two stimuli taken separately yield significantly different values of $\bar{\theta}$, then the curve of conditioning to a compound of the two stimuli should be expected to deviate further than either of the separate curves from a simple growth function. The only relevant experiment we have discovered in the literature is one reported by Miller (16); Miller's results appear to be in line with this analysis, but we would hesitate to regard this aspect of the theory as substantiated until additional relevant data become available.

Although we shall not develop the argument in mathematical detail in the present paper, it may be noted that the statistical association theory yields certain specific predictions concerning the effects of past learning upon the course of learning in a new situation. In general, the increment or decrement in p during any trial depends to a certain extent upon the immediately preceding sequence of trials. Suppose that we have two identical animals each of which has $p(n)$ equal, say, to 0.5 at the end of trial n of an experiment, and suppose that for each animal response A is reinforced on trial $n+1$. The histories of the two animals are presumed to differ in that the first animal has arrived at $p(n)=0.5$ via a sequence of reinforced trials while the second animal has arrived at this value via a sequence of unreinforced trials. On trial $n+1$, the second animal will receive the greater increment to p (except in the equal θ case); the reason is, in brief, that for both animals the stimulus elements most likely to occur on trial $n+1$ are those with high θ values; for the first animal these elements will have occurred frequently during the immediately preceding sequence of trials and thus will tend to be preponderantly connected to A prior to trial $n+1$; in the case of the second animal, the high θ elements will have been connected to $\bar{A}$ during the immediately preceding sequence and thus when A is reinforced on trial $n+1$, the second animal will receive the greater increment in weight of connected elements. From this analysis it follows that, other things equal, a curve of reconditioning will approach its asymptote more rapidly than the curve of original conditioning unless extinction has actually been carried to zero. How important the role of the unequal θ distribution will prove to be in accounting for empirical phenomena of relearning cannot be adequately judged

until further research has provided means for estimating the orders of magnitude of the effects we have mentioned here.

SUMMARY

Earlier statistical treatments of simple associative learning have been refined and generalized by analyzing the stimulus concept in greater detail than heretofore and by taking account of the fact that different components of a stimulating situation may have different probabilities of affecting behavior.

The population of stimulus events corresponding to an independent experimental variable is represented in the statistical model by a mathematical set. The relative frequencies with which various aspects of the stimulus variable affect behavior in a given experiment are represented by set operations and functions.

The statistical model, taken together with an assumption of association by contiguity, provides a limited theory of certain conditioning phenomena. Within this theory it has been possible to distinguish aspects of the learning process that depend upon properties of the stimulating situation from those that do not. Certain general predictions from the theory concerning acquisition, extinction, and relearning, are compared with experimental findings.

Salient characteristics of the model elaborated here are compared with other quantitative formulations of learning.

REFERENCES

1. BROGDEN, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
2. BUSH, R. R., & MOSTELLER, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.
3. BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
4. CALVIN, J. S. Decremental factors in conditioned-response learning. Unpublished Ph.D. thesis, Yale Univer., 1939.
5. CRAMÉR, H. *Mathematical methods of statistics*. Princeton: Princeton Univer. Press, 1946.
6. ESTES, W. K. Effects of competing reactions on the conditioning curve for bar pressing. *J. exp. Psychol.*, 1950, **40**, 200–205.
7. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
8. FELLER, W. *An introduction to probability theory and its applications*. New York: Wiley, 1950.
9. GRANT, D. A., & HAKE, H. W. Dark adaption and the Humphreys random reinforcement phenomenon in human eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 417–423.
10. GULLIKSEN, H., & WOLFLE, D. L. A theory of learning and transfer. *Psychometrika*, 1938, **3**, 127–149.
11. GUTHRIE, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, **43**, 1–20.
12. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
13. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
14. HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
15. MILLER, G. A., & MCGILL, W. J. A statistical description of verbal learning. *Psychometrika*, in press.
16. MILLER, J. The rate of conditioning of human subjects to single and multiple conditioned stimuli. *J. gen. Psychol.*, 1939, **20**, 399–408.
17. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.
18. THURSTONE, L. L. The learning function. *J. gen. Psychol.*, 1930, **3**, 469–493.

[MS. received November 12, 1952]

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Volume 66, 1952

STUDIES IN REPUTATION: I. SEX AND GRADE DIFFERENCES IN SCHOOL CHILDREN'S EVALUATIONS OF THEIR PEERS. II. THE DIAGNOSIS OF SOCIAL ADJUSTMENT. Read D. Tuddenham. #333. $1.50.

RELIABILITY AND VALIDITY OF JUDGES' RATINGS OF ADJUSTMENT ON THE RORSCHACH. Marguerite Q. Grant, Virginia Ives, and Jane H. Ranzoni. #334. $1.00.

SOME PERSONALITY CORRELATES OF RELIGIOUS ATTITUDES, AS DETERMINED BY PROJECTIVE TECHNIQUES. Ralph Mason Dreger. #335. $1.00.

PERSONAL SECURITY AS RELATED TO STATION IN LIFE. Andie L. Knutson. #336. $1.00.

THE VALIDITY OF PERSONALITY-TRAIT RATINGS BASED ON PROJECTIVE TECHNIQUES. Henry Samuels. #337. $1.00.

RORSCHACH ESTIMATES OF PERSONALITY ATTRIBUTES IN THE MICHIGAN ASSESSMENT PROJECT. Woodrow Wilbert Morris. #338. $1.00.

DEVELOPMENT OF AN EYE CAMERA FOR USE WITH MOTION PICTURES. Paul R. Wendt. #339. $1.00.

FACTOR ANALYSIS OF SOME REASONING TESTS. Harold M. Carter. #340. $1.00.

THE USE OF VOCATIONAL INTEREST SCALES IN PLANNING A MEDICAL CAREER. Edward K. Strong, Jr., and Anthony C. Tucker. #341. $2.00.

THE OCCUPATIONAL LEVEL SCALE AS A MEASURE OF DRIVE. Gordon J. Barnett, Irving Handelsman, Lawrence H. Stewart, and Donald E. Super. #342. $1.50.

OBJECTIVE MEASUREMENT OF REALITY-CONTACT WEAKNESS. Albert Kreinheder. #343. $1.00.

THE PERFORMANCE OF SCHIZOPHRENIC AND NORMAL INDIVIDUALS FOLLOWING FRUSTRATION. Harold Wilensky. #344. $1.00.

PARENTS' ATTITUDES VS. ADOLESCENT HOSTILITY IN THE DETERMINATION OF ADOLESCENT SOCIOPOLITICAL ATTITUDES. Kenneth Helfant. #345. $1.00.

PERSONALITY OF THE CONVULSIVE PATIENT IN MILITARY SERVICE. T. W. Richards. #346. $1.00.

THE ACADEMIC ACHIEVEMENT OF VETERAN AND NONVETERAN STUDENTS. Norman Frederiksen and W. B. Schrader. #347. $1.50.

THE PRIMARY PERSONAL VALUES MEASURED BY THE ALLPORT-VERNON TEST, "A STUDY OF VALUES." Hubert E. Brogden. #348. $1.00.

SELECTED PERSONALITY VARIABLES AND THE LEARNING PROCESS. Eugene L. Gaier. #349. $1.00.

JUDGING INTERESTS FROM EXPRESSIVE BEHAVIOR. N. L. Gage. #350. $1.00.

Orders for any of these Monographs can be placed separately at the prices listed above, or the entire volume can be ordered for $6.00.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1333 Sixteenth Street N. W.

Washington 6, D. C.

Ready for Fall Classes . . .

METHOD AND THEORY IN

Experimental Psychology

By CHARLES E. OSGOOD

Associate Professor of Psychology and Research Associate, Institute of Communications Research, University of Illinois

This book covers a major portion of the field of experimental psychology, and represents an important synthesis of fact and critical theory. It includes a wide range of material on perception, learning, sensory and symbolic processes. In addition, it evaluates and integrates all of the principal contemporary theories, with special emphasis on clarifying the issues between the gestalt and behavioristic schools.

Method and Theory in Experimental Psychology is intended chiefly for undergraduate majors, graduate students and professional psychologists. Its penetrating theoretical insights and wealth of experimental evidence make it a significant and long-needed contribution to the field. *With 223 illustrations and 21 tables.*

Approximately 976 pages Fall, 1953 Probable price, $10.00

OXFORD UNIVERSITY PRESS 114 Fifth Avenue, New York 11, N.Y.